



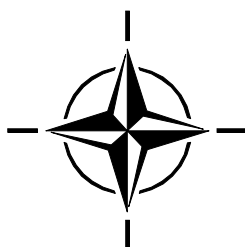
STO TECHNICAL REPORT

TR-HFM-185

Processes for Assessing Outcomes of Multi-national Missions

(Processus d'évaluation des résultats
de missions multinationales)

Findings of Task Group HFM-185.



Published November 2013

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE NOV 2013		2. REPORT TYPE N/A		3. DATES COVERED	
4. TITLE AND SUBTITLE Processes for Assessing Outcomes of Multi-national Missions				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Science and Technology Organization North Atlantic Treaty Organization BP 25, F-92201 Neuilly-sur-Seine Cedex, France				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Multi-national missions require comprehensive assessment that measures the impact of military and civilian actions on opposing forces and the local population. In order to identify and characterize limitations of current assessments and identify best practices, this NATO RTG task group examined current measurement and assessment practices by NATO members in current and recent theatres of operations. The task group also conducted a preliminary study to identify a potential set of core indicators that could provide a high-level assessment of macro change over time to compliment campaign assessments aimed at specific interventions. These core indicators, if used to assess all missions, would allow for cross-mission comparison, a capability that currently does not exist. Based on the findings, the task group recommends that NATO develop the policies, plans, infrastructure and procedures to successfully conduct comprehensive assessments. In order to do this NATO must develop standard assessment procedures and frameworks for use throughout NATO/ISAF and other multi-national operations. This includes common lexicons and definitions used by member nations and a core set of metrics to supplement campaign measures of performance (MOPs) and measures of effectiveness (MOEs) and allow for comparisons between NATO missions.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 78	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			





STO TECHNICAL REPORT

TR-HFM-185

Processes for Assessing Outcomes of Multi-national Missions

(Processus d'évaluation des résultats
de missions multinationales)

Findings of Task Group HFM-185.

The NATO Science and Technology Organization

Science & Technology (S&T) in the NATO context is defined as the selective and rigorous generation and application of state-of-the-art, validated knowledge for defence and security purposes. S&T activities embrace scientific research, technology development, transition, application and field-testing, experimentation and a range of related scientific activities that include systems engineering, operational research and analysis, synthesis, integration and validation of knowledge derived through the scientific method.

In NATO, S&T is addressed using different business models, namely a collaborative business model where NATO provides a forum where NATO Nations and partner Nations elect to use their national resources to define, conduct and promote cooperative research and information exchange, and secondly an in-house delivery business model where S&T activities are conducted in a NATO dedicated executive body, having its own personnel, capabilities and infrastructure.

The mission of the NATO Science & Technology Organization (STO) is to help position the Nations' and NATO's S&T investments as a strategic enabler of the knowledge and technology advantage for the defence and security posture of NATO Nations and partner Nations, by conducting and promoting S&T activities that augment and leverage the capabilities and programmes of the Alliance, of the NATO Nations and the partner Nations, in support of NATO's objectives, and contributing to NATO's ability to enable and influence security and defence related capability development and threat mitigation in NATO Nations and partner Nations, in accordance with NATO policies.

The total spectrum of this collaborative effort is addressed by six Technical Panels who manage a wide range of scientific research activities, a Group specialising in modelling and simulation, plus a Committee dedicated to supporting the information management needs of the organization.

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS System Analysis and Studies Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These Panels and Group are the power-house of the collaborative model and are made up of national representatives as well as recognised world-class scientists, engineers and information specialists. In addition to providing critical technical oversight, they also provide a communication link to military users and other NATO bodies.

The scientific and technological work is carried out by Technical Teams, created under one or more of these eight bodies, for specific research activities which have a defined duration. These research activities can take a variety of forms, including Task Groups, Workshops, Symposia, Specialists' Meetings, Lecture Series and Technical Courses

The content of this publication has been reproduced directly from material supplied by STO or the authors.

Published November 2013

Copyright © STO/NATO 2013
All Rights Reserved

ISBN 978-92-837-0180-4

Single copies of this publication or of a part of it may be made for individual use only by those organisations or individuals in NATO Nations defined by the limitation notice printed on the front cover. The approval of the STO Information Management Systems Branch is required for more than one copy to be made or an extract included in another publication. Requests to do so should be sent to the address on the back cover.

Table of Contents

	Page
List of Figures/Tables	v
List of Acronyms	vi
Acknowledgements	viii
HFM-185 Membership List	ix
Executive Summary and Synthèse	ES-1
Chapter 1 - Introduction	1-1
1.1 Rationale and Background	1-1
Chapter 2 - Assessment and Measurement	2-1
2.1 Introduction	2-1
2.2 Assessment Explained	2-1
2.3 Why Conduct Assessments?	2-2
2.4 Good Assessment	2-2
2.5 Bad Assessment	2-3
2.6 Measurement Explained	2-3
2.7 Good Measurement	2-3
2.8 NATO Orientation	2-5
Chapter 3 - Analyst, Commander, and Staff Interviews	3-1
3.1 Introduction	3-1
3.2 Methods and Approach	3-2
3.2.1 Swedish Research	3-3
3.2.2 Dutch Research	3-4
3.2.3 British Research	3-4
3.2.4 Canadian Research	3-4
3.2.5 American Research	3-5
3.3 Different Militaries and their Assessment Structures	3-5
3.3.1 Swedish Deployment of Analysts	3-5
3.3.2 Dutch Deployment of Analysts	3-5
3.3.3 British Deployment of Analysts	3-5
3.3.4 Canadian Deployment of Analysts	3-6
3.3.5 American Deployment of Analysts	3-6

3.4	Key Findings	3-6
3.4.1	Lack of Standard Assessment Procedures and Measures	3-8
3.4.2	Transparent and Accessible Knowledge and Data Management	3-11
3.4.3	Linkages between Planning and Assessment	3-12
3.5	Conclusions	3-15
Chapter 4 - Assessments by Non-Military Actors		4-1
4.1	Introduction	4-1
4.2	World Health Organisation (WHO)	4-1
4.2.1	WHO Statistical Information System (WHOSIS)	4-1
4.2.2	Global Health Atlas	4-2
4.3	Freedom House (FH)	4-4
4.3.1	Databases	4-4
4.4	United Nations Educational, Scientific and Cultural Organisation (UNESCO)	4-4
4.4.1	Institute for Statistics	4-5
4.5	World Bank	4-5
Chapter 5 - Core Variables to Assess Change over Time		5-1
5.1	Approach	5-1
5.2	The Set of Indicators	5-1
5.3	Demonstration	5-3
5.4	Conclusions	5-7
Chapter 6 - Conclusions and Recommendations		6-1
6.1	Introduction	6-1
6.2	Replicable Assessment Standards and/or Data Framework	6-1
6.3	Common Lexicon and/or Terminology	6-2
6.4	A Set of Core Variables	6-3
6.5	Integrating Assessment and Planning	6-3
6.6	Consistent Training in Assessment	6-4
6.7	Information Technology and Data Management	6-4
6.8	Strong Leadership of the Assessment Process	6-5
6.9	Recommendations	6-5
6.9.1	Measurement and Assessment Process Recommendations	6-5
6.9.2	Operational Process Recommendations	6-5
6.9.3	Future Research Recommendations	6-5
Chapter 7 - References		7-1
ANNEX A - US Interview Protocol		A-1

List of Figures/Tables

Figure		Page
Figure 3-1	British Deployment Approach	3-11
Figure 3-2	Connecting Assessments and Planning	3-14
Figure 4-1	Example of a Static Map	4-3
Figure 5-1	Male and Female Employment Rates as a Function of Time	5-7
Tables		
Table 2-1	Typically Accepted Values for the S.M.A.R.T. Keyword	2-4
Table 3-1	Rank of Respondent Interviewed as a Function of Country	3-3
Table 3-2	Summary of Themes and Enablers of Good Assessment	3-7
Table 4-1	Examples of Relevant Measures for Standard of Life, Comparing the Latest Available data (November 12, 2009) for Congo and the Netherlands, Found in WHOSIS	4-2
Table 5-1	Set of Indicators and SPSS Variable Names Based on World Bank Data	5-2
Table 5-2	Descriptive Statistics for the Demonstration Dataset	5-3
Table 5-3	Total Variance Explained for Demonstration Dataset	5-4
Table 5-4	Rotated Component Matrix for Demonstration Dataset	5-5
Table 5-5	Correlations between Year and Other Variables	5-6

List of Acronyms

AAG	Afghan Assessments Group
ACT	Allied Command Transformation
BDA	Battle Damage Assessments
CEA	Campaign Effects Assessment
CEFCOM	Canadian Expeditionary Force Command
CIDNE	Combined Information Data Network Exchange
COIN	Counter-Insurgency
COS	Chief of Staff
DAC	Development Assistance Committee
DALY	Disability-Adjusted Life Year
DSTL	Defence Science and Technology Laboratory
EBAO	Effects Based Approaches to Operations
FH	Freedom House
GBD	Global Burden of Disease
HEA	Human Effects Advisor
HFM	Human Factors and Medicine
HQ	Headquarter
IDC	Information Dominance Centre
IED	Improvised Explosive Device
IJC	ISAF Joint Command
IPM	Integrated Police training Mission
ISAF	International Security Assistant Force
JEAC	Joint Effects Assessment Cell
MNF-I	Multi-National Force - Iraq
MoD	Ministry of Defence
MOE	Measure of Effectiveness
MOP	Measure of Performance
MPICE	Measuring Progress In Conflict Environments
NATO	North Atlantic Treaty Organization
NLD	Netherlands
OA	Operations Analyst
OECD	Organisation for Economic Cooperation and Development
OPLAN	Operational Plan
ORSA	Operations Research and Systems Analyst
PRT	Provincial Reconstruction Teams

RC	Regional Command
RIPTOA	Relief In Place/Transfer Of Authority
RTG	Research Task Group
RTO	Research Technology Organisation
SIGACTS	Significant Activities
TFH	Task Force Helmand
TFU	Task Force Uruzgan
UK	United Kingdom
UNESCO	United Nations Educational Scientific and Cultural Organisation
UIS	UNESCO Institute for Statistics
US	United States
WHO	World Health Organisation
WHOSIS	World Health Organisation Statistical Information System

Acknowledgements

The members of NATO RTO Task Group (HFM-185) would like to thank Dr. Barbara Sotirin (USA) and Dr. Peter Essens (NLD) for their initiative, planning and effort for starting the Research Task Group (RTG) and for providing initial guidance to the Group. We would also like to thank Dr. Allison Astorino-Courtois and Ms. Tessa Baker from NSI, Inc. for assisting the Task Group by conducting interviews with producers and consumers of assessments, analysing the data, and reporting the results. Their very high calibre work and dedication to the project were valuable assets to the project. As well, we extend special gratitude and appreciation to Dr. Peter Essens for his peer review of the document and his excellent mentorship and guidance.

HFM-185 Membership List

Katherine M. Banko, PhD

Behavioral Psychologist
Defence Research and Development Canada
Centre for Operational Research and Analysis
Department of National Defence
101 Colonel By Drive
Ottawa, ON K1A 0K2
CANADA
Phone: +1 (613) 945-2388
Fax: +1 (613) 992-3342

Katherine.banko@forces.gc.ca

Peter Berggren, PhD

Cognitive Psychologist
Swedish Defence Research Agency
Division of Information and Aeronautical Systems
Department of Man
P.O. Box 1165
SE-58111 Linköping
SWEDEN
Phone: +46 13 37 8222
Fax +46 13 12 49 38

Peter.berggren@foi.se

Aletta R. Eikelboom, MSc.

Reserve Officer, Major
Consultant
TNO Behavioural and Societal Sciences
Kampweg 5
3769 DE Soesterberg
P.O. Box 23
NETHERLANDS
Phone: +31 888 666 006

Aletta.eikelboom@tno.nl

Ivy V. Estabrooke, PhD

Thrust Manager, Human Social Cultural and Behavioral Sciences
Expeditionary Maneuver Warfare and Combating Terrorism Department, Code 30
Office of Naval Research
UNITED STATES

Ivy.estabrooke@navy.mil

Trevor Howard

[Dstl] Principal Analyst S2O
PCS Department
iSAT N, Room C010
Grenville West Court
Dstl Portsdown West
Portsdown Hill Road
Fareham
PO17 6AD
UNITED KINGDOM
Tel: +44 (0)2392 53
Fax: +44 (0)2392 53 2348

Trhoward@dstl.gov.uk

Roman Lau, Dipl.Ing.

Lieutenant Colonel (GS)
COIN Network
German Armed Forces Command & Staff College
Manteuffelstraße 20
22587 Hamburg
GERMANY
Phone: +49 (40) 8667-5250
Fax: +49 (40) 8667-6209

romanlau@bundeswehr.org

Jenny Marklund

Research Psychologist
Swedish Defence Research Agency, FOI
Division for Defence Analysis
Department of Peace Support Operations
SE 16440 Stockholm
SWEDEN
Phone: +46 709 27 74 76
Fax: +46 8 555 031 00

Jenny.marklund@foi.se

Maj. Joakim Marklund

Staff Officer CD&E
Swedish Armed Forces
Headquarters
Joint Operations J3
Section for Strategic Operations
Lindingövägen 24
107 85 Stockholm
SWEDEN
Phone: +46 8 7887500
Fax: +46 8 788 77 78

Joakim.marklund@mil.se

Elmer L. Roman

Oversight Executive
Office of the Secretary of Defense
Acquisition, Technology & Logistics
Research and Engineering
1225 South Clark Street
Arlington, VA 22202
UNITED STATES
Phone: 703-607-5019
Fax: 703-601-2370

Elmer.roman@osd.mil

A.J. (Tony) van Vliet, PhD

Senior Researcher (Psychologist)
TNO Behavioral and Societal Sciences
Kampweg 5
3769 DE, Soesterberg
NETHERLANDS
Telephone: +31 88 866 59 36

Tony.vanvliet@tno.nl

Andy Williams

Section Head - Solutions Analysis
NATO HQ Supreme Allied Commander Transformation HQ SACT
7857 Blandy Rd, Suite 100
Norfolk, VA 23551-2490
UNITED STATES
Phone: +1 (757) 747-3876
Fax: +1 (757) 747-3271

Andrew.williams@act.nato.int



Processes for Assessing Outcomes of Multi-national Missions

(STO-TR-HFM-185)

Executive Summary

Military forces, including NATO, have expanded the breadth of Battle Damage Assessments (BDA) from addressing the impact of kinetic actions in terms of target hits and casualties, to a more comprehensive assessment that measures the impact of military and civilian government and non-government actions (“whole of government” or “comprehensive approach”). Assessment in this context is described as the use of data to inform and modify judgement about the current situation and recent trends and is used for informing the planning process, the results of which can provide evidence of the impact of previous activities that is useful for informing new plans.

This NATO RTO Task Group (HFM-185) examined current measurement and assessment practices by its participating Nations in current and recent theatres of operations in order to identify and characterize limitations of current assessments and therefore identify best practices. The Task Group also conducted a preliminary study to identify a potential set of core indicators that would be able to assess macro change in a region over time and allow for meaningful comparisons of results between missions. Finally, the Task Group provided a set of recommendations to improve assessment for multi-national missions. As part of this research, interviews were conducted with individuals who either performed assessments or were the consumers of assessment products. Participants included civilian and military members of the Swedish, Dutch, Canadian, British and American militaries/departments of defence.

Results demonstrated that there is high diversity within and between nations on concepts, approaches and quality, highly driven by the individual commander. We concluded that there is a high need for well-defined and consistent terminology, standard assessment practices that link assessments to the planning process, and a requirement for core measures that are consistent across mission assessments to allow for comparisons between missions outcomes, and yet complimentary to specific campaign intervention assessments. An enabler for proper assessment is a transparent and accessible data and knowledge management system. In order to identify a small comprehensive set of variables that were observable, culture-free, counted the same way by everyone and independent, the Task Group conducted a systematic and comprehensive review of databases and datasets of public organisations, such as World Bank, World Health Organisation. After multiple subject matter experts reviewed and assessed the variables based on specific criteria, a set of 33 variables that described seven principal components; demographics, economy, education, health, infrastructure, law/security and social constructs, were identified. A (statistical) demonstration of the ability of these indicators to show macro change is provided. It was concluded that these datasets can serve as well-grounded indicators of the long-term development of a region and allow comparison between missions. For this to happen, commanders should understand the rationale for collecting data for the core variables in order to support these efforts in their own assessments.

The HFM-185 Task Group recommends that:

- 1) NATO develop the policies, plans, infrastructure and procedures to successfully conduct and harmonise comprehensive assessments.

- 2) NATO must develop standard assessment procedures and frameworks for use throughout NATO/ISAF and other multi-national operations. This includes common lexicons and definitions used by member nations and a core set of metrics to supplement campaign Measures of Performance (MOPs) and Measures of Effectiveness (MOEs).
- 3) Data/knowledge and information system should be developed to provide data at the lowest level of classification to NATO members and partners.
- 4) Assessments must be integrated into the standard procedures for development of campaign plans.
- 5) Senior leadership take ownership of the assessment process and assure there is adequate training of relevant personnel on the standardized assessment approach or framework.
- 6) Further refinement and validation of a core variables set should be addressed in a follow-on RTG to support harmonisation of assessment efforts.

Processus d'évaluation des résultats de missions multi-nationales (STO-TR-HFM-185)

Synthèse

Les forces militaires, y compris l'OTAN, ont élargi le champ des évaluations des dommages de combat (BDA, *Battle Damage Assessment*), qui s'étend maintenant de l'analyse d'impact des actions cinétiques en termes de coups au but et de pertes jusqu'à une évaluation plus complète mesurant l'impact des actions gouvernementales et non gouvernementales civiles et militaires (« approche holistique » ou « approche globale »). Dans ce contexte, une évaluation est définie comme une utilisation des données destinée à informer et modifier un jugement sur la situation actuelle et les tendances récentes. L'évaluation sert à renseigner le processus de planification, dont les résultats peuvent apporter des preuves sur l'impact des activités précédentes, preuves utiles au renseignement de nouveaux projets.

Le groupe de travail de la RTO de l'OTAN (HFM-185) a examiné les pratiques actuelles de mesure et d'évaluation employées par les pays participants dans les théâtres d'opérations récents et présents afin d'identifier et caractériser les limites des évaluations actuelles et identifier ainsi les meilleures pratiques. Le groupe de travail a également mené une étude préliminaire pour identifier un ensemble d'indicateurs centraux permettant d'évaluer au fil du temps les changements macroscopiques d'une région et de comparer de manière constructive les résultats de différentes missions. Enfin, le groupe de travail a émis un ensemble de recommandations visant à améliorer l'évaluation des missions multi-nationales. Dans le cadre de cette étude, des entretiens ont été réalisés avec des personnes chargées des évaluations ou utilisant le produit des évaluations. Les participants ont inclus des membres civils et militaires des armées / ministères de la défense suédois, hollandais, canadien, britannique et américain.

Les résultats ont démontré l'existence d'une forte diversité au sein des pays et entre eux sur les concepts, approches et qualité, en grande partie du fait du commandant. Nous avons conclu qu'il était fortement nécessaire d'instaurer une terminologie et des pratiques d'évaluation standard bien définies et cohérentes, qui relient les évaluations au processus de planification. Il convient également d'instaurer des mesures centrales communes à toutes les évaluations de mission – afin de pouvoir comparer les résultats de différentes missions – et qui complètent cependant les évaluations spécifiques d'interventions en campagne. Un système accessible et transparent de gestion des données et connaissances faciliterait l'évaluation. Afin d'identifier un jeu complet et restreint de variables observables, non influencées par une culture, indépendantes et comptées de la même manière par chacun, le groupe de travail a examiné intégralement et avec méthode les bases et ensembles de données d'organisations publiques comme la Banque mondiale ou l'Organisation mondiale de la santé. Après analyse et évaluation des variables à l'aide de critères spécifiques, plusieurs experts en la matière ont identifié un ensemble de 33 variables détaillant sept composantes principales (démographie, économie, éducation, santé, infrastructure, droit/sécurité et constructions sociales). Une démonstration (statistique) de la capacité de ces indicateurs à signaler les changements macroscopiques est fournie. Il en a été conclu que ces ensembles de données peuvent servir d'indicateurs pertinents du développement à long terme d'une région et ainsi permettre de comparer différentes missions. Pour ce faire, les commandants d'opérations doivent comprendre la raison d'être du recueil des données relatives aux variables centrales afin de soutenir ces efforts dans le cadre de leurs propres évaluations.

Le groupe de travail HFM-185 recommande que :

- 1) L'OTAN développe les politiques, projets, infrastructures et procédures visant à mener et harmoniser des évaluations complètes.
- 2) L'OTAN développe des cadres et procédures d'évaluation standards pour une utilisation au sein de toute l'OTAN / la FIAS et des autres opérations multi-nationales. Cela comprend des lexiques et définitions communs utilisés par les pays membres et un ensemble central d'indicateurs complétant les Mesures de Performance (MOP) et Mesures d'Efficacité (MOE) en champagne.
- 3) Des systèmes d'informations et de données/connaissances soient développés pour fournir des données au plus bas niveau de classification aux membres et partenaires de l'OTAN.
- 4) Les évaluations soient intégrées aux procédures standards de développement des plans de champagne.
- 5) Les hauts dirigeants prennent eux-mêmes la responsabilité du processus d'évaluation et veillent à ce que le personnel concerné reçoive une formation appropriée sur le cadre ou l'approche d'évaluation normalisée.
- 6) Le perfectionnement et la validation plus poussée d'un ensemble de variables centrales doivent être traités lors d'un RTG ultérieur afin de soutenir l'harmonisation des efforts d'évaluation.

Chapter 1 – INTRODUCTION

Membership of the North Atlantic Treaty Organization Research and Technology Organization Human Factors and Medicine-185 Research Task Group (NATO RTO HFM-185 RTG) examining “Processes for Assessing Outcomes of Multi-national Missions” includes Canada, Germany, Netherlands, NATO Allied Command Transformation (ACT), Sweden, United Kingdom and United States. The Group is led by Canada.¹

1.1 RATIONALE AND BACKGROUND

In the past, most military assessments were conducted as Battle Damage Assessments (BDA) primarily reflecting the number of target hits, casualties, and deaths [1]. By necessity, today’s assessments must employ a wider lens and measure not only the impact of military activities on opposing forces, but also the impact of military and civilian actions on local populations. This shift in focus to Effects Based Approaches to Operations (EBAO) coincided with NATO operations in Bosnia-Herzegovina and Kosovo and the counterinsurgency (COIN) and counterterrorism objectives of multi-national forces in Afghanistan and Iraq, which require assessments more comprehensive than simply the number of insurgent forces killed or captured [2].

Measuring Progress in Conflict Environments (MPICE) was one identified framework developed by the US Institute of Peace, US Army Corps of Engineers, and US Army Peace Keeping and Stability Operations Institute to systematise and co-ordinate cross-regional data collection and assessment of social, economic, and political progress in stabilisation and reconstruction operations [3]. It was piloted during humanitarian and disaster relief efforts with reports of some success [4]. To prolong the findings, it was suggested that the framework be extended for use in counter-insurgency operations, specifically in Afghanistan. The emphasis of the framework was on measuring outcomes, with hundreds of items (goals, indicators, and measures) and having the agility to measure and track progress from imposed stability operations to a state of self-sustaining peace. Despite these two intended strengths, analyses of the framework revealed that it contained several flaws and inconsistencies [5].

When the Exploratory Team handed over to what would become the HFM-185 Task Group in 2008, it stipulated five specific objectives for the Group. These objectives were to:

- 1) Research and develop a common understanding of current assessment practices for whole of government activities;
- 2) Identify NATO requirements for an assessment framework;
- 3) Investigate feasibility/applicability of MPICE and other frameworks;
- 4) Adapt and experiment in a field trial the most applicable framework against NATO requirements; and
- 5) Formulate recommendations for refinement and implementation of the framework.

In 2009, the Team began to examine the feasibility of validating MPICE for use by NATO. The MPICE framework represented a substantive effort in advancing the systematic collection and coordination of cross-regional data collection in order to assess social, economic, and political progress in Stabilization and

¹ The Task Group was originally co-lead by Netherlands and United States. However, due to career, the US lead had to resign from the Panel. The Canadian member, after consultation and acceptance from the Group, assumed the role and responsibility of Chair.

INTRODUCTION

Reconstruction Operations. Because of these strengths, it seemed likely that the MPICE framework could be adapted for use by NATO. While other assessment frameworks, other than MPICE, were available they could not be effectively transitioned for use in NATO, because NATO is limited to military actions [for example see 6, 7].

Through careful presentation and discussion of member-nations' assessment practices, it became clear that each member nation conducted assessments in nation-specific ways, often undertaking several different types of assessment in a given conflict/stability environment (i.e., campaign assessments, intelligence assessments). As such, the Team determined that in order to resolve some of the issues involved in developing an assessment framework that would be adopted by NATO Nations required foremost a solid understanding of the current state of assessment practices applied in theatres of conflict and stability operations. These changes resulted in a shift of focus in the overall objectives guiding the Task Group.

In order to develop a common understanding, the Task Group decided to research current member nation practices for multi-national missions' assessments to identify key limitations and challenges of conducting assessments in conflict environments and contexts, as well as distil best practices for multi-national assessments. One initial observation of this review of current practices was the frequently changing set of metrics employed by military commanders in conflict environments. In addition, assessments typically conducted for commanders tended to focus on tactical level campaign assessments and therefore were unable to either assess change at a global level or see how their efforts impacted at that level. Consequently, the Group sought to identify a small set of core indicators that reflected global change which could be used to compliment campaign assessments and allow for comparisons across multi-national missions, a comparability that is currently not straightforwardly possible.

The Task Group focussed on:

- Examining current measurement and assessment practices in theatres of operations and identify key limitations and challenges and distil best practices;
- Identifying a set of core indicators; and
- Producing recommendations to improve assessment of multi-national missions.

In Chapter 2, assessment, measurement and their relationship are briefly explained. Chapter 3 presents the research identifying current assessment practices. Chapter 4 introduces assessment practices and available data sets by non-military organisations in order to identify what has been achieved in assessing mission areas, country development and/or societal change. This will provide a starting point to discover indicators that may be useful in assessing global aspects of multi-national operations. Chapter 5 presents the method for core indicator selection followed by statistical analyses of how they relate and are useful for high level assessment that compliments campaign assessments.² Chapter 6 provides conclusions and recommendations based on the overall findings of this multi-faceted study.

² The detailed technical analyses included in this chapter are intended for those familiar with advanced statistical knowledge.

Chapter 2 – ASSESSMENT AND MEASUREMENT

“However beautiful the strategy, you should occasionally look at the results”

Winston Churchill

2.1 INTRODUCTION

A basic finding of the Task Group’s research was the lack of consistent definitions of “assessment” and “measurement” across member-nations.¹ This Chapter seeks to explicate definitions of assessment and demonstrate the difference between “good” and “bad” assessment and measurement. The authors of this report recognise that there are several operational factors contributing to assessment challenges, including:

- Operations are complex and turbulent, making it difficult to establish reliable baselines or to distinguish between strategic shifts in the environment and shorter term fluctuations [8];
- Command structures are complex with multiple reporting lines and many layers and key stakeholders distributed across many organisations [9];
- There are often multiple actors [local, national, and international] undertaking a range of concurrent activities with different underlying aims and objectives over different time horizons [10];
- Interventions are often politicised which creates different pressures in measuring progress and can lead to perceived imperatives to achieving tangible outputs [4];
- The unpredictable and rapidly evolving nature of operations requires regular reframing of intervention logic, testing of assumptions, and adaptation of activities [11];
- Change is often very hard to measure, and cause and effect is very difficult to establish [12];
- Identifying and collecting suitable indicators of change to develop a composite picture requires a variety of metrics, both qualitative and quantitative, drawing from subjective and objective data [13]; and
- There are many practical difficulties in data collection and evaluation, including the security of data collectors and the reliability of data sources [14].

2.2 ASSESSMENT EXPLAINED

The word ‘assessment’ has multiple uses and meanings. In the context of this report, assessment is the use of data to inform and modify judgement about the current situation and recent trends. In addition to past events, assessment can include known future events, plans, intentions, actions and assumptions to help extrapolate current trends to the future in order to develop a forecast of the future situation. The objective of an assessment is to synthesise insights regarding the current situation to provide feedback that influences the decision maker’s behaviour.

The NATO COPD (Comprehensive Operational Planning Directive) [15] and the NATO Operations Assessment Handbook² [16] proposes a definition of [operations] assessment as:

‘the activity that enables the measurement of progress and results of operations in a military context, and the subsequent development of conclusions and recommendations that support decision making’.

¹ At the time of this research, there were two NATO reports underway: Comprehensive Operational Planning Directive (COPD) [10] and the Operations Assessment Handbook Version 1.0 [5]. Both reports are in draft status.

² NATO Operations Assessment Handbook, Interim Version 1.0, dated 29 January 2011.

2.3 WHY CONDUCT ASSESSMENTS?

CAVEAT: There are many reasons to conduct and use assessment, each use depends on member-nations' policies regarding the dissemination of information, which may be mission critical.

Assessments should lead to improved planning. The process of determining how to measure achievement of an effect or objective (discussed further below) enhances conceptual understanding and should lead to better designed plans and more insightful objectives [15].

Assessment conclusions and recommendations feed adjustments into the plan and give an evidential basis for decision making. At the strategic level, assessments may measure progress and results of activities, situations and organisations outside of NATO military plans, for the purpose of strategic decision making [15].

Results of assessment and ensuing recommendations allow for a more informed allocation of resources and/or funds to those areas that need it most [17].

Operations assessment produces substantial material for the identification of best practices for whole of government activities, and for the historical study of operations contributing to the development of lessons identified and increased corporate knowledge [18].

Planning involves the setting of targets, and should give the commander specific and measurable targets at which to aim the efforts of his staff and forces, in concert with non-military actors. Assessment can be a means to motivate by confirming and enabling the celebration of success when achieved, or highlighting and dealing with failures, which can also be useful for public information and media briefing purposes.

2.4 GOOD ASSESSMENT

Assessments must remain true to its objective. The objective of an operation assessment is to produce insights pertaining to the current situation and to provide feedback to support decisions by the leadership.

Assessment must take a multi-dimensional perspective. It is essential to build the assessment by looking at the environment through multiple perspectives that cross lines of operations and time periods.

Assessment teams must never compromise the integrity of their reports. They need to be free to express their findings about the current conditions and the influential factors they discover. The assessment must not be manipulated to produce 'good news'.

Good assessment requires the assessment team to have access to a wide array of information and people in order to perform their job properly.

The activities related to intelligence, knowledge management and assessments are very similar, thus generating the potential for confusion or duplication of effort. Good assessment will be aided by a mutually supporting relationship between these activities. There will be a shared understanding of this symbiotic relationship.

Good assessment requires a conceptual framework to integrate various data streams into an analysable or interpretable whole. Illustrating basic relationships between factors and then operationalizing those factors to the specific geographic scope of inquiry involves understanding the levels of analysis and the levels to which you can aggregate validly.

2.5 BAD ASSESSMENT

Bad assessment can generally be characterised by an absence of those things that make up good assessment. Bad assessment can also come about through using poor analytical techniques or poor data.

Examples are:

- Small sample sizes and inappropriate sample frames;
- Absence of systematic measurement variables;
- Abstracting data from different levels of analysis (different resolution/granularity);
- Issues related to data aggregation; and
- Discontinuities in measurement approaches without appropriate methodological acknowledgements.

Additional information about good and bad assessment is available in the NATO Operations Assessment Handbook [15], the SAS-091 Final Report [19], Assessment Process for RC(SW) [20].

2.6 MEASUREMENT EXPLAINED

Measurement is the evaluation of the amount, size, or extent of something. In operations assessment there are two generally used terms:

- Measures of Performance (MOPs); and
- Measures of Effectiveness (MOEs).

An MOP is a criterion used to assess friendly actions tied to measuring task accomplishment.³ MOPs measure the extent of progress in execution of the plan (are the actions being executed as planned?)

An MOP can measure either quantitatively or qualitatively the amount of effort that is put [input indicator] into an activity or it can measure the output [output indicator] of an activity:

- Input indicators are criteria used to assess the level of effort being applied in terms of resources (e.g., manpower or funding).
- Output indicators are criteria used to assess the direct products of activities (e.g., number of patrols conducted, number of wells dug, quality of government officials).

An MOE is a criterion used to assess changes in system behaviour, capability, or operational environment that is tied to measuring the progress of a mission, achievement of an objective, or creation of an effect. They give an indication of the success of the plan and results achieved.

2.7 GOOD MEASUREMENT

Good measurement should be readily observable, culturally independent, and able to be aggregated without anything more than an average or a sum (no indices).

³ There is a general lack of agreement regarding the specific definitions of MOPs and MOEs across NATO and ISAF member-nations. For a full consideration of the definitions of MOEs and MOPs see RTO-TR-HFM-160, "How to Improve your Aim: Measuring Effectiveness of Activities that Influence Attitudes and Behaviors" [21].

ASSESSMENT AND MEASUREMENT

It is naturally much easier to measure progress towards a goal if that goal is clearly expressed. Objectives should be **S.M.A.R.T.** as shown in Table 2-1[21].⁴

Table 2-1: Typically Accepted Values for the S.M.A.R.T. Keyword.

Letter	Major Term	Minor Terms
S	Specific	Significant, Stretching, Simple
M	Measurable	Meaningful, Motivational, Manageable
A	Agreed	Attainable, Actionable, Achievable, Appropriate, Aligned, Acceptable
R	Realistic	Results-oriented, Relevant, Resourced, Resonant
T	Time-bound	Timely, Trackable, Tangible, Time-oriented, Time-limited

Regardless how careful the measurement process is, it is essential to exercise great care in the selection of suitable and appropriate MOEs for each of the effects. Limited time, resources and availability of reliable data sources can hamper this process and often only after a period of data gathering, can the more useful MOE be identified and selected for further analysis.

MOEs must always be mission specific, be sensitive to change (i.e., increases or decreases in behaviours that indicate progress and/or failure) and, in addition, should be:

- Mission Related;
 - Mapped to a desired objective and/or effect;
 - Meaningful;
 - Measurable;
 - Culturally and Locally Relevant;
 - Time and Geographically bounded⁵;
 - Be at an appropriate level of analysis;
 - Cost/Time Effective (in terms of manpower).
- } Essential
- } Highly desirable

There should be a set of consistent measures which should be invariant over the duration of the campaign. This allows for systematic measurement over time which permits the identification of trends and anomalies.

With this appreciation for what constitutes good and bad assessment, the HFM-185 team members sought to explore existing practice in the field. The following Chapter relates findings from assessment producers and consumers regarding their experience in assessing MOEs in Afghanistan and other current multi-national missions.

⁴ There is no clear consensus on which five keywords should be used. Choosing certain combinations can cause duplication, such as “attainable” and “realistic”. Therefore, the team chose the ones that resonated across the multiple country membership.

⁵ It should be stated which geographic region the MOE is being measured on, for example, by Province, by Region or by entire Area of Operation (AOO).

2.8 NATO ORIENTATION

Although the importance of assessment and MOE is already emphasised in NATO documentation [15, 16], there is still a need for analytically and scientifically derived and practically usable indicators/MOE. Planners and decision makers in NATO Operations still are not provided with practical orientation concerning such specific variables.



Chapter 3 – ANALYST, COMMANDER, AND STAFF INTERVIEWS

3.1 INTRODUCTION

The purpose of this Chapter is to describe findings from experiences of assessment practices in the field. The Task Group surveyed commanders, operations analysts, and members of military staffs regarding the production and use of comprehensive assessment in multi-national operations. The results from a series of in-depth interviews and workshops conducted with American, British, Canadian, Dutch and Swedish veterans of multi-national operations are presented herein. The Chapter concludes by drawing out the common themes and providing recommendations for improved assessment practices. The objectives of the interview protocol attached as Annex A included:

- Discussing commanders' and practitioners' current national and NATO assessment practices in theatre and the key limitations to conducting these;
- Identifying the key features of assessment practices both in the organisational and procedural sense; and,
- Identifying improvements in assessment processes, training and education, leadership, and other emergent themes.

The interviews used the working NATO definition at that time, which was¹:

“The purpose of assessment is to support the NATO political domain² and Commander and staff decision making in three areas:

- 1) The progress of plan execution (actions / tasks),
- 2) The effectiveness of those executed actions by measuring the achievement of results (effects, objectives, and the end-state),
- 3) Drawing conclusions about past situations, and in some cases mak[ing] forward-looking estimates about future trends, and recommend[ing] adjustments to the plan based on these conclusions.”

Two items about the conduct of this study should be noted. First, while a standardised interview protocol was followed, due to the variation in interviewees' nationality and operational experience, the balance of interview questions were open-ended³ to allow researchers latitude to explore different perspectives on comprehensive assessments. In addition, researchers had the freedom to modify the interview protocol to match the unique circumstances of their research subjects. For example, some of the Dutch information and all of the British data was collected during workshops that were not designed with this research project in mind, but still provided usable content and information. The qualitative interview data collected was from a total sample of 70 respondents.

¹ The definition of assessment changed in March 2011 to “Operations Assessment is to be understood as the activity that enables the measurement of progress and results of operations in a military context, and the subsequent development of conclusion and recommendations that support decision making” (p. 1-1).

² Research team participants struggled with how to define and characterize NATO's “political domain,” for the purposes of this report, the team concluded that the political domain included those non-kinetic aspects of multi-national operations, including humanitarian and governance assistance and activities. This terminology was removed from the NATO definition released on March 2011 in the NATO Operations Assessment Handbook.

³ Open-ended questions are those to which a respondent is allowed to give the response or information that seems most appropriate to him or her in a free form manner. This is in contrast to closed-ended survey or interview questions which can be answered with a set of discrete responses like “yes” or “no”, or with a specific piece of information (e.g., age range).

Second, the majority of interviews were based on respondents' recent experiences with NATO operations in Afghanistan (International Security Assistance Forces (ISAF)) and/or Iraq (Multi-national Force-Iraq (MNF-I)). Although a few respondents also had experience in Bosnia/Kosovo, the balance of what is reported in this Chapter is based on MNF-I and ISAF activities.

We believe that a "recent experience bias" does not diminish the value of interviewees' insights as it is reasonable to expect that future NATO missions will require population-centric assessment to support stability operations, and/or development and governance assistance.

The following sections report the Task Group's findings regarding current comprehensive assessment practices including what interviewees identified as enablers of good assessment. The first section provides a detailed discussion of the research design, methods, and approach.

3.2 METHODS AND APPROACH

As noted above, interviewees included two types of individuals, those primarily responsible for producing "comprehensive" assessments and those who mostly consumed assessments. Researchers did not initially seek to generalise across nations. Whilst each interview was conducted individually, researchers often referenced prior discussion with other respondents in order to gather either confirmatory or contradictory evidence. The research team aimed to diversify the sample across nationality, rank, operational experience and mission location. Interviewees were recruited from Canada, Sweden, the United Kingdom and the United States and interviews or workshops were conducted by nationals of interviewees' own countries. As a consequence of the sheer variety of individual respondents and their own experiences, 'assessment' came to mean different things to different respondents. Nonetheless, interviews, but not necessarily workshops, were framed around the NATO definition of 'comprehensive assessment' to provide a consistent frame of reference. Despite this, it is not generally clear that all respondents had the same type of assessment in mind, which becomes one of our chief findings. The following section details the approach and methods adopted by the country research teams.

INFORMATION SOURCES: Researchers conducted interviews with respondents in their home countries: Canada, Sweden, the Netherlands and the United States and workshops with respondents in the Netherlands and the United Kingdom. The breakdown of the information streams is as follows:

- Fifteen interviews with Swedish personnel (14 military, 1 civilian) with experience from deployment to Regional Command (RC) North and/or Provincial Reconstruction Teams (PRT) Mazar-e-Sharif [22];
- Two interviews with a Dutch (NLD) Operations Analyst and Commander with experience involving the Task Force Uruzgan, as well as a compiled list of best practices and lessons learned based upon the collective Dutch experience in Uruzgan from 2006-2010 derived from a workshop in which 12 Operational Analysts participated, all with experience serving with Task Force Uruzgan;
- Nine surveys administered among Canadian Operational Analysts and Human Effects Advisors (HEAs) with experience in Afghanistan and Bahrain;
- Thirteen commander and staff interviews among American officers and 2 Canadian generals with experience in both Afghanistan and Iraq; and
- A 2008 workshop with 19 British civilian operations analysts that sought to "evaluate the UK 'frontline' operational analysis community's position on Campaign Effects Assessment (CEA), based on their experience up to the end of 2008."

The ranks and nationalities of participants are shown in Table 3-1.

Table 3-1: Rank of Respondent Interviewed as a Function of Country.

Rank	SWE	NLD	GBR	CAN	USA	Total
Lieutenant	1					1
Captain		8			1	9
Major	7	5		2		14
Lt. Colonel	3				4	7
Colonel	3	1			2 ⁴	6
General				2	4	4
Civilian	1		19	7	2	29
TOTAL	15	14	19	9	13	70

CAVEAT ABOUT GENERALISABILITY⁵: While researchers sought to employ the same or similar interview protocols across all interviews, deviations were both necessary and expected in order to reflect the unique national experiences and native languages of each respondent. As a consequence, care should be taken when generalising from these results. Though the results cannot be strictly generalised due to a small sample size and the qualitative nature of the interview protocol, the approach used elicited the necessary feedback to enable extraction of general themes and a consequent summary of best practices for comprehensive assessments.

3.2.1 Swedish Research

The Swedish research involved 14 male and one female respondents. Twelve of the Swedish participants had worked in the Swedish-led PRT between 2008 and 2010 and three had worked in Regional Command (RC) North headquarters. The male participants were Army officers ranging in rank from lieutenant to colonel. The female respondent was a civilian development advisor. The Swedish interviews sought to probe military practitioners' understanding and the application of assessments of progress within their PRTs. The interviewees were selected based on their experiences from working in key positions with responsibility for planning and assessments. Interviewees were selected from personnel who had recently worked in Afghanistan (between 2008 and 2010). The following positions were represented at the PRT-level: Commanding Officer, Chief of Staff, Assisting Chief of Staff G2⁶, Assisting Chief of Staff G3, Deputy Chief G5 and Staff Officer G5. At RC North interviewees represented: Chief CJ5, Staff Officer CJ5 and Staff Officer CJ3.

Swedish research participants were divided into two sub-groups. Sub-group 1 participants (March 2010) were asked to discuss current assessment input and guidelines, the assessment in practice, and feedback to planning.

⁴ Including one US Navy Captain

⁵ The concept of generalizability refers to the extent to which research results and conclusions drawn from one case or data sample can be validly applied to cases beyond those included in the study. The issue of the generalizability of results is a key issue in any study where anything less than an entire population has been studied. As an example, consider one of the most common examples: polling. The question, say in a poll of Afghan adults' opinions of the Afghan Government, would be whether the population sample polled (e.g., adult residents of Kabul) is representative enough of the total population of Afghan adults that the results and conclusions produced could reasonably be used to draw conclusions about that larger population.

⁶ Under the continental staff system, the "G" code is used to refer to general staff.

Sub-group 2 participants (September 2010) were asked about interaction and collaboration issues including the dissemination of operational objectives, collaboration on the assessment(s); and recommended improvements.

All interviews were semi-structured in nature and lasted 45 - 90 minutes depending on how much time the participants were able to set aside for the interview. Two researchers participated in each interview, where one was responsible for asking the questions and the other one for taking extensive notes. Interviews were recorded, but were not transcribed. To ensure that the researchers had captured the content of the interview correctly all notes were sent to the interviewees, giving them a chance to clarify any misunderstandings. The Swedish study is reported in greater detail in the report, “*Challenges in Assessing Progress in Multifunctional Operations: Experiences from a PRT in Afghanistan*” [22] - all information regarding the Swedish national experience contained in this Chapter is drawn from this research.

3.2.2 Dutch Research

Dutch input is based on experiences with operational analysis and comprehensive assessment during Netherlands’ ISAF mission in Uruzgan, Afghanistan (2006-2010). Respondents included participants from an Operations Analyst workshop and two additional interviews. Twelve experienced operational analysts participated in the one-day workshop. The workshop aimed to identify best practices and lessons learned from four years of operational analyst deployment in Task Force Uruzgan. All operational analysts are reserve officers in the rank Major and Captain and have at least one time been deployed to Uruzgan. The workshop focused on several topics including: linking planning to assessment, operational analysts in long-term and short-term planning, and data collection vs. analysis. The interviewees were a Dutch Colonel and a Captain. The Colonel served as the commander of the Task Force Uruzgan and the Captain as an operational analyst. The interviews were semi-structured and lasted approximately 90 minutes.

3.2.3 British Research

British input to this Chapter is largely taken from a report that details information collected during a two day workshop hosted by UK Defence Science and Technology Laboratory (Dstl) in September 2008 among practitioners of CEA within the operational analyst teams in the United Kingdom and UK-lead headquarters and commands [23]. Most of the 19 operational analysts in attendance had more than 20 years of experience and discussed a range of issues, including identifying best practices whenever possible. In Britain, all operational analysts are civilian government employees.

3.2.4 Canadian Research

Due to time, number of researchers and geographical constraints on the Canadian research ‘team’, the standard interview protocol was converted into a questionnaire. The questionnaire was emailed to 11 persons with deployment experience, nine of whom participated. The question format was a series of open-ended questions that closely mirrored the interview protocol. The questionnaire was sent out in early December 2010; responses were requested within a four-week period. All research participants received an email regarding the nature of the research, the expected time required to complete the survey, and how the data would be used. Respondents were informed that their names would be removed from the documents and the information would be handed over to fellow researchers who would conduct the analysis; the data would only be reported in aggregate form. These two processes were put in place to ensure participants confidentiality and anonymity.

Despite the different method employed for Canadian participants, open-ended responses and feedback were consistent with the findings from Sweden, the Netherlands, the United Kingdom and the United States.

3.2.5 American Research

All participants fell into at least one of the following categories: military commanders; senior-level civilian advisors or leadership; leader of assessment team, cell, or staff; military assessment personnel; civilian operational analysts; operational personnel who work with assessment results; and civilian personnel involved in assessments; however, most had recent experience in multiple roles.

The US research group, comprised of three trained researchers, rotated their roles as primary interviewer or note-taker. A snowball sampling technique was employed with US respondents where after a first round of contacts, candidate interviewees were suggested by participating interviewees. The researchers also conducted two interviews with Canadian Generals with experience as RC South and Taskforce Kandahar Commanders. Interviews were conducted from January 2011 - April 2011. Three interviews were conducted in person; the remaining ten were conducted over the phone.

Written transcripts were produced for each interview and were reviewed for correctness and completeness by the interviewee before being considered final. Audio recordings were used only for in-person interviews and internal transcription purposes.

3.3 DIFFERENT MILITARIES AND THEIR ASSESSMENT STRUCTURES

3.3.1 Swedish Deployment of Analysts

Sweden's OAs (Operational Analysts) are civilians, however Sweden does not currently deploy operations analysts to Afghanistan. Rather, OAs are posted and trained for service at higher Swedish headquarters like Permanent Joint HQ, Force HQ, and Brigade staffs.

3.3.2 Dutch Deployment of Analysts

Operational analysis is a fairly new capability of the Royal Dutch Army. Since the start of the NLD mission in Uruzgan (RC(S)) (2006), the Netherlands has supported its HQ with a team of two Operational Analysts. All Task Force Uruzgan (TFU) HQs between 2006 and 2010 were supported by operational analysts. Currently the Netherlands Integrated Police training Mission (IPM) in Kunduz (RC(N)) is also supported by two analysts. In the TFU, the OA was officially placed within the G5 branch and in the IPM the OA was placed directly under the Chief of Staff. Most analysts were deployed for 4 months, and most teams had a staggered rotation, meaning one analyst rotated in when the other was halfway through his/her tour. Due to lack of capacity, not all positions were always filled (this especially was the case for the IPM in Kunduz).

All of the Dutch OAs were reserve officers in the rank Captain or Major and most have a background in social science and/or operations research. There are currently 18 OAs, organised in a network. Although a training program is still in development, all OAs have received military and OA training.

3.3.3 British Deployment of Analysts

The United Kingdom supports each of its HQ with a team of Operational Analysts (usually two but sometimes three, four, or a singleton post). Currently in Afghanistan, HQ Task Force Helmand (TFH) is supported by two analysts. One would deploy and return with the Brigade HQ on its 6 month tour; the other would deploy approximately halfway through the tour and return halfway through the next Brigade HQ's tour, thus providing continuity through the relief in place. All of the analysts are Ministry of Defence (MoD) civilians with a background in Operational Analysis.⁷

⁷ Operational Analysis is the term used in the UK military for Operations Research.

ANALYST, COMMANDER, AND STAFF INTERVIEWS

The lead analyst in the team or a singleton post will be an analyst with experience of supporting deployed formation HQ. Where more than one analyst is deployed, they will, where possible, have complementary expertise such as pairing a social scientist with a more technically focused analyst.

All analysts are deployed through the Support to Operations Group of the Dstl. Each analyst is given individually targeted training prior to deployment to bring them to a common baseline. Whilst deployed analysts are expected to be generalists rather than specialists, they will each have their own specialisation and, through Dstl, have reachback into the entirety of the UK analytical and scientific community.

3.3.4 Canadian Deployment of Analysts

Canada's defence scientists are civilians and deployed with relevant task forces and teams. There is no uniform training for analysts, with respect to assessment measures and practices; analysts received various types of training (e.g., statistical refresher course) if it was available and if there was time to accommodate it prior to deployment. Regardless, analysts came prepared with a variety of skills and training. Despite discipline (e.g., psychologist versus physicist), all found numerous ways to provide useful analyses and assessments. Typically one or two analysts deploy with a taskforce and are embedded with the J5 branch responsible for long-term plans and assessments. Defence scientists have also provided (and continue to provide) assessment support in addition to that of the various task forces deployed in Kandahar province. Tours last between four months and one year depending on the individual circumstances of the analyst. All analysts have access to reach back support from the larger defence scientific community.

3.3.5 American Deployment of Analysts

For the non-US NATO audience, it is important to note that within the US military, OAs typically are uniformed officers of the military, and known as Operations Research and Systems Analysts (ORSA). According to one interviewee, there were fewer than 15 ORSAs deployed to Afghanistan during his service there (2009 - 2010). Indeed, the interviewee in question had to volunteer his own analytic services to facilitate the work of the ISAF Afghan Assessments Group. The small number or absence of trained analysts forces commanders responsible for submitting assessments to higher echelons to adapt and develop their own assessment approaches. Thus, the US often relies upon its PRT Commanders to coordinate and conduct assessments in their area of responsibility (AOR) without any formal training in measurement theory or quantitative methods. The lack of formal assessment training and defined roles for analysts in theatre creates inconsistency in terms of the assessments produced by the US military and across the coalition/Alliance. Despite differences in the organisation of assessments in different member nations, countless areas of convergence were found (i.e., common themes).

3.4 KEY FINDINGS

The following section discusses interview results, and is organised along three areas that emerged from the interview data:

- 1) Data Frameworks;
- 2) Knowledge Management; and
- 3) Planning and Assessments. Common themes and assessment practice improvements recommended by interviewees are also discussed. Table 3-2 presents a summary of the themes and enablers of the results.

Table 3-2: Summary of Themes and Enablers of Good Assessment.

Major Theme	Enabler(s)
Replicable Assessment Standards/ Data Framework	<ul style="list-style-type: none"> • Standard assessment procedures • A core set of metrics • Consistent training in assessment prior to deployment • Common lexicon/terminology
Data Management and Knowledge Management	<ul style="list-style-type: none"> • Accessible and integrated information system at the lowest possible clearance level • Facilitation of better transition of assessment practices during Relief in Place/Transfer of Authority (RIPTOA)
Integrating Assessments and Planning	<ul style="list-style-type: none"> • Integration of assessments into the campaign planning process • Coordination and integration civilian and military assessment activities • Strong leadership of the assessment process

In the latter stages of the Afghanistan and Iraq operations, comprehensive assessments sought to measure the effects of ISAF's/MNF-I's COIN strategy which shifted the operational focus from defeating a combatant force to defending and protecting a local population.⁸ Consequently, assessments generally developed into population-centric approaches in the same way that the operational plan was gauged to securing the population and improving their autonomy. Taken in the aggregate, the interviews reveal some overarching themes surprisingly common across nationality, rank and role (i.e, assessment producer or consumer). However, while this section enumerates these common themes the wide variation in the ranks and skill sets of individuals tasked with conducting assessments across country makes it difficult to develop universal recommendations for structural refinements within NATO-member militaries.

A key finding is that there exist several gaps in the substance and practice of producing assessments. The first of these gaps or deficits occurred in the course of conducting this study: at present ***there exists no common understanding or consistent terminology regarding assessment practices or even what constitutes "comprehensive assessment."*** This was found to be the case both among Alliance members and within their own governments. The comparable British term is "comprehensive effects assessment," which is codified in British doctrine as "evaluation of campaign progress, based on levels of subjective and objective measurement in order to inform decision-making" [23]. US and Dutch respondents for example were unfamiliar with the term "comprehensive assessment" as used in this context. Not only were assessments "lost in translation" across units of different nationalities, but issues were often "lost in transcription" between unit members from the military and civilians attached to Provincial Reconstruction Teams (PRTs).

COMMON THEMES: In considering the interview data as a whole the most surprising finding was that rather than identifying "best practices" for producing and disseminating comprehensive assessments, respondents

⁸ See ISAF Commander's Counterinsurgency Guidance http://www.nato.int/isaf/docu/official_texts/counterinsurgency_guidance.pdf; also the National Military Strategy, 2004 <http://www.defense.gov/news/Mar2005/d20050318nms.pdf>.

overwhelmingly cited gaps, deficits and where improvements were needed with regard to assessments of this type. **In fact, the majority of respondents were unable to identify any particularly useful, well-done and/or valid comprehensive assessments that they either produced or consumed.** Consequently, overall results are more in the order of lessons learned and suggested improvements than best practices as was the initial intent of this study.

The common concerns that did emerge were of three general types:

- a) The ***lack of standard assessment procedures and core measures*** resulting in unreliable, and at times inconsistent and personality-driven assessments, which hinder comparable measurement of perceived mission progress across geography, time, and troop rotations;
- b) The need for ***transparent and widely accessible knowledge and data management***; so that knowledge reliably gained is not lost across geography, time, and troop rotations; and
- c) The ***lack of practices that link assessment staffs with planning activities***, while also ensuring that both military and non-military assessments are used as part of the planning process, and vice versa. This includes both consistent procedures to link operations planning and assessments, as well as practices that help to integrate military and non-military efforts across a whole of government approach. The development of procedures to better link planning and assessments within the military as well as across government might help to generate more coordinated “whole of government,” cross-government, or inter-departmental plans that include Lines of Operation (LOO) like development and governance.

In many cases, these themes pointed directly to the need for enhanced or additional enablers; where applicable, suggested enablers are indicated in discussions of each of these themes.

3.4.1 Lack of Standard Assessment Procedures and Core Measures

Regardless of country, most interview respondents noted the absence of consistent, replicable assessment frameworks to inform their own data collection efforts or those of their staff. Ideally, frameworks would provide guidance on how to measure the (strategic) objectives through a path that leads from objective to MOEs/MOPs. Such metrics would ideally be incorporated into the Intelligence Collection Plan, which allows for multiple assets, like unit patrols or trainers, to be tasked to frequently report on the metrics. Many interviewees believed that assessment taskings were often *ad hoc* in nature with varying timelines, had different underlying objectives with inconsistent or no operational metrics at all. Moreover, interviewees reported that there appeared to be a lack of benchmarks and baselines against which assessments could be compared, and metrics changed so frequently that trend analysis was almost impossible. The critical information that respondents felt was missing in both requests to produce assessments, and as a consequence, in the actual assessments produced were:

- What are we measuring against (i.e., baselines/benchmarks, objectives)?
- What is the specific definition of progress/regress vis-à-vis plan objectives?
- How do we actually measure, or operationalise, that change reliably and accurately in the field?
- How should multiple information streams be collated to inform a thorough comprehensive assessment?

Several respondents noted that headquarters (HQ) and regional command level commanders knew what the objectives of a given operational plan and its associated assessments were, but that these objectives were not broadly or clearly disseminated to lower echelons. Thus, PRT commanders were often left to develop their own assessment frameworks in an *ad hoc* fashion that sought to explore their own tactical concerns as well as the more strategic ambitions of the campaign plan. The lack of clear assessments guidelines and frequently changing timelines created difficulty at lower echelons as PRTs endeavoured to carry out the

“Clear, Hold, Build” strategy, while also addressing disparate information requests. Respondents in both the American and Swedish studies nearly universally noted that “the guidelines of what was to be assessed were vague and left much room for personal interpretation” [22]; thereby creating the perception of “subjectivity” in many assessments. The UK report also notes that in Afghanistan it was evident “that in the design, development and application of Campaign Effects Assessment (CEA) there is little communication between the formations. Even when comparing the CEA processes between each of the four Regional Commands in Afghanistan, there is little parity in the approaches” [24].

Interestingly, HQ and strategic-level interviewees tended to think plans and assessments were relatively coordinated with clear timelines and update schedules. However, as noted above, this was not necessarily the perception of more junior and operational personnel.

Respondents also remarked on the need to have an understanding of the operational environment prior to conducting assessments. As one American General noted:

“[In order to do a comprehensive assessment], you have to understand and identify the underlying problem of what you are going to assess. Before we did a fundamental theatre assessment, we conducted what General McChrystal called a “listening tour.”

A number of respondents were frustrated by the perceived lack of standardised assessment metrics and measures across regional commands and rotations which they felt left assessments unnecessarily vulnerable to individual biases, or even, as one senior officer noted, to political “manipulations” within staffs. In short, the lack of standards allowed reports to be overly personality driven. As one Canadian respondent noted:

“Part of this situation extends from the degree to which personalities drive the mission and approaches such as that embodied in the paraphrased words of Task Force Kandahar Headquarters 5-10 Chief of Staff (TFK HQ 5-10 COS) indicate: Every boot on the ground collects the information necessary to conduct assessments. However, these “boots” conducted this process with no consistency or understanding of what was entailed, and the information was very loosely consolidated and controlled. Subsequently, any assessments derived from them were subject to subjective input and command veto. That is, if the assessment derived from the reports contradicted those of someone more senior they were subject to amendment.”

Given a high operational tempo, the predominance of the military in most multi-national, whole of government operations, the emphasis on force protection, and the relative ease of accessing significant activities and events data versus other types of data, respondents frequently reported what they perceived as an undue emphasis on kinetic and security operations over other lines of operation. Such a focus on security combined with inconsistent data integration techniques diminishes the reliability of assessments intended to be comprehensive (i.e., including events other than blue and red force activities).

Since the connections between different data sectors (i.e., governance, security, development, strategic communications) is not fully understood in terms of the broader goal of stability (in Afghanistan), many assessment producers were left to develop their own weighting schemes. The British report similarly noted the inappropriate use of weighting schemes and the lack of subject-matter expertise. Because the development of weighting schema often suffered from implicit cognitive biases, the most important contributors to stability often varied significantly across the same area of operation. Respondents consistently noted that it was “difficult to interpret the results and determine the relevance of the information depending on varying credibility.” Each commander or OA was often on his own to determine which information to trust, which information to discard, and which information to weight most heavily in reaching a conclusion. According to interviewees, this absence of a common framework for conducting assessments, especially in Afghanistan, resulted in a methods vacuum which was filled by individual and often competing assessment approaches.

ANALYST, COMMANDER, AND STAFF INTERVIEWS

Respondents were also unaware of a common foundational data aggregation framework or approach. They were also unclear about what constitutes good assessment in Afghanistan (and Iraq). Oftentimes the data that was most readily or cheaply available, in terms of both money and manpower, formed the basis of assessments, rather than data that reflected specifically determined metrics that matched the campaign plan and objectives. As J. Marklund, et al. noted “instead they assessed what was possible to measure, for example, number of events or activities” rather than what might be truly meaningful or informative to the commander or assessments consumer [22].

In order to remedy these limitations, several interviewees highlighted the need to create a ***standardised assessment framework and process*** that is adopted across the Alliance and train the ***relevant assessors throughout each military and civilian organisation*** operating in the multi-national mission, planners and assessors could achieve an improved synergy that would address some of the plan-assessment issues discussed further below. A standardised process for developing relevant indicators with recommended measurement methodologies would help individuals on the ground to determine what to assess and would address the observed lack of standardisation necessary for comparative analyses. In addition, such a standardised approach (one example of which is described in Chapter 6), would allow for the implementation of theory-based methodologies and remedy the *ad hoc* nature of assessment approaches particularly identified in Afghanistan [22]. As one Canadian Human Effects Advisor noted, “This is a situation that needs to be rectified and standardised at the strategic level [A common assessment system] needs to be adopted at the strategic level and then ample training and exposure provided to subsequent users. If it is to be a legitimate piece of the operational process it must also be aligned with planning and resource management.”

Development of a consistent, standardised assessment methodology is likely to be difficult, but the development of this methodology is critical to resolving other issues identified during the course of this research, including the need for a knowledge management system that meshes well with the assessment protocol and the need for training. One Canadian respondent noted that:

“Despite efforts to develop a common framework [for assessment], everyone’s information requirements and objectives seemed to be different. This was driven by a number of factors:

- 1) The constantly changing cast of characters. For instance, when he arrived in Iraq, as Commanding General MFN-I, General Petraeus changed the criteria for measuring success. Subsequently, his Joint Effects Assessments Cell (JEAC) had to redesign the tool(s) they were using to track the success of the plan. In this sense, it appears that planning is not truly tied to effects, rather the effects that are measured and how that measuring is done is driven more by the personal in charge.
- 2) Secondly, the demands of different agencies drove the need for different responses.
- 3) Lastly, the plan itself changed sufficiently with each new command and/or HQ that the conduct of assessments needed to shift to keep pace.”

There was almost universal support among respondents for such a common approach including a standard set of terminology.

As one US Marine with experience in Afghanistan noted; the development of a common lexicon would be an improvement compared to the current situation. “Comprehensive assessment is a term that I never heard there [in Afghanistan], which is not to say that we didn’t do it.” In fact, none of the Americans interviewed were familiar with “comprehensive assessments.” Such standards, a common terminology, and a transparent process would improve the legitimacy of assessment findings according to one American interviewee serving in a reach-back capacity who noted that “Ensuring the process for developing assessments is credible and transparent will alleviate some of the concerns from higher headquarters on the legitimacy and relevance of any given assessment.”

In addition, respondents indicated a similar level of support for the idea of developing assessment standards for metrics, reporting and the use of assessment tools by which everyone abides; otherwise, respondents were concerned by the perceived subjectivity and manipulability of existing assessments approaches. Regardless of the assessment framework, however, respondents were very supportive of increased training to ensure that data collectors were equipped with a basic understanding of common collection protocols; that individuals tasked with conducting analysis were able to recognise reliable data sources and statistics; and that assessors were able to collate and combine incoming data streams into a replicable overall assessment. Particularly among Americans, it was very apparent that there often inadequate social science expertise among assessment staff in Iraq and especially in Afghanistan.

Increasing analysts' dwell time in theatre and increasing the amount of overlap time between incoming and outgoing analysts would improve assessments by providing analysts with greater familiarity with the operational theatre and an opportunity to learn from their predecessors in the accepted assessment protocol. As an American respondent noted, "... we need planners and operators who are familiar with the challenges of assessing and are familiar with the sorts of things that have been identified as indicators of desired effect".

British brigades in Afghanistan have a well-developed handover process. Each brigade HQ will visit theatre several times during its pre-deployment preparation. A series of work-up exercises are run (identical for each brigade as far as possible, reflecting the situation in theatre) at which officers from the most recently returned brigade act as mentors. Operational analysts attend these exercises to train with the HQ with which they will be deployed (see Figure 3-1). The Netherlands has a comparable preparation and handover process, and the OAs also rotate out of sync. Although the staggered rotation is optimal to support information exchange, one of the disadvantages is that one of the OA's is deployed between two HQ and will therefore not entirely be seen as part of that HQ.

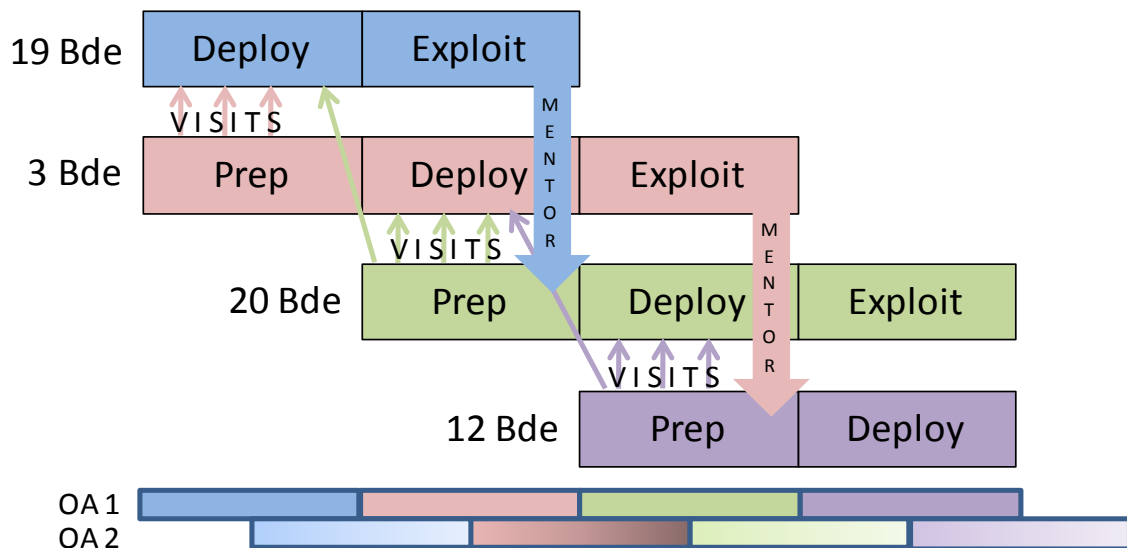


Figure 3-1: British Deployment Approach.

3.4.2 Transparent and Accessible Knowledge and Data Management

Interviewees with experience of conducting or storing assessments noted that there was inadequate information technology and knowledge management infrastructure to effectively leverage information and ensure that it remained accessible and could be shared with current operators in theatre, those in reach-back cells, and future rotations of commanders and assessments staff. As one Canadian HEA noted:

“There is a huge amount of data stored in several locations within the HQ; however, because there is no common system of storage and the storage systems are being constantly changed or amended it can be both difficult to find and even more difficult to understand as to its origins and intent”.

Other US respondents noted the need to ensure that comprehensive assessments are “discoverable,” and that data is extractable from assessments products to facilitate its use by others and eliminate redundant and unnecessary data collection efforts.

Likewise, Dutch OAs involved in the Netherlands TFU recognised the importance of such an infrastructure to facilitate consistency across rotations into and out of theatre, particularly in terms of ensuring that the “measurement instruments and methods used by the previous OA team is handed over correctly and complete(ly) to the next OA team.” The absence of such an infrastructure as well as standardized procedures (discussed further below) often forced individuals tasked with assessment to reinvent the wheel of measurement of success and assessments during each new rotation - a finding repeated consistently across country and researcher. The communication of results across different networks and clearance levels was also difficult. As a Dutch interviewee noted, “it was difficult to communicate data between different data networks, e.g. to communicate the results of assessments to others/next higher level because these parties were often on different networks with other classification status.”

While each unit and team frequently had access to information technology, centralized servers, and the Combined Information Data Network Exchange (CIDNE), no system-wide data and information sharing procedures or tools were in place to ensure that data would be accessible across deployments and over time. For instance, as one Dutch respondent noted, “assessment data and report(s) were stored, however, passwords sometimes prevented use of older documents (the passwords were forgotten or nobody knew how to access old documents); sometimes documents disappeared because of hardware failure; backups could not be made to take home.”

Access to an *effective and integrated knowledge management system* would greatly improve the utility of assessments information and encourage further cross-pollination across units that currently limit their observations to within their own stovepipe (mostly security). As the British research team noted, “the military must undergo a cultural change to foster and invest in good information management.” Better knowledge management and IT might facilitate the use of reach-back support cells in units’ home countries, which could be “useful for conducting standardised analyses and technical support”. This knowledge management and information technology infrastructure should be made available at the lowest possible information restriction and classification level to also allow for sharing of information contained in the system with local population leadership (like leaders of the ANSF and local governors/mayors). Such information sharing might help to improve the sustainability of assessments and the transition to indigenous control in Afghanistan and elsewhere.

3.4.3 Linkages between Planning and Assessment

Interviewees indicated that there often is a weak link between operational planning and assessment activities. As noted, PRT commanders were often uncertain how to operationalize and measure many of the Operational Plan’s (OPLAN) objectives. While OPLANs often included specific goals for development, security, and governance (the major Lines of Operation), they rarely, if ever, included specific operationalizations of each goal - leaving assessment producers to their own devices. Likewise, it was often difficult for commanders and assessment producers to maintain a consistent set of core metrics due to the frequency of additional information requests emanating from commands.

According to the Swedish study, respondents found it difficult to interpret the appropriate time scales and extent from the ISAF OPLAN. In addition, there was little use of assessments to review the current operational plan of the PRT in Mazar-e-Sharif. RC North often required weekly assessments without

“providing usable assessment tools” [22]. Yet, Swedish research participants believed that assessments had some effect on decisions regarding ongoing operations. This disconnection between assessments and their perceived role in revising operational plans often fostered a sense of discontent regarding the utility of some assessments products. As one American noted, “...I never received any further guidance based on the reporting we gave [in Afghanistan]. I don’t know if they had any resources to attack the problem”.

Many American interviewees noted the multiplicity of assessments that PRT Commanders and their staffs are required to feed. According to one American interviewee, “the Afghan Assessments Group (AAG) does a monthly, quarterly, and year-long assessment which does not match up with the 6-week ISAF Joint Command Information Dominance Center (IJC/IDC) district assessment”. Most PRT commanders, however, only revised their campaign plan at the beginning of their rotation and the ISAF campaign plan is only typically revised annually. By better aligning existing information requirements and the timescales therein, the demand on assessments producers can be minimized and redundancies eliminated.

Most respondents suggested that assessments could be greatly improved by better coordination in planning and execution between the civilian and military branches (“whole of government”). As the Swedish researchers noted, “the Swedish-Finnish PRT ...has to harmonise both civil and military means in stability and reconstruction operations, counterinsurgency operations, and peace support operations [22].” Participants noted that “the military and SIDA (Swedish International Development Agency) had been sent to the PRT with different mandates, objectives and cultures, without practical instructions on how to cooperate and coordinate [22].” While the plan naturally comes before any assessment scheme designed to analyse its efficacy, it is important that there be a close connection between any plan and its assessment strategy - as the plan evolves, so should the assessment approach.

Regardless of nationality, interviewees noted the “tunnel - visioned” nature of assessments that often focused on the more military security activities over other lines of operations. Most interviewees indicated that measures of the security of ISAF were the most readily available metrics with some information available regarding population security. However, assessments in any kind of stability operation will naturally be population-centric, meaning that any assessment that does not branch out beyond security measures is providing an incomplete picture of the operations space.

An ***integrated assessment and campaign plan*** that provides recommendations for operations, including the best way to measure given plan elements, while specifying designated reporting intervals that are consistent across the entire theatre of operations is a critical enabler of good assessment. Improved assessment also requires better ***coordination and integration*** of civilian development and diplomatic agencies into the assessments process, in order to harness shared information and expertise. To ensure quality assessment, it is necessary for a single voice/senior leader to exert strong influence over the methods, content, scheduling, and documentation of reporting. As one Canadian respondent noted, “assessment needs to be coordinated at the highest level from which it will be sought from the operational team, in this instance Canadian Expeditionary Force Command (CEFCOM)” regarding the importance of assessment. An American interviewee reinforced this view when he noted that in Afghanistan, “they really needed leadership to come in and establish the needs, who was going to acquire the information, etc., to allow for the production of a fairly consistent report that can produce monthly, comparable reports.”

Figure 3-2, which resulted from a Dutch workshop among OAs, depicts how the link between planning and assessment might be realised. The left column shows the different levels of planning in line with the design of long-term comprehensive missions, OPLANS are developed, including short term (e.g., 6 months) objectives and effects. In order to contribute to the OPLANS (or other plans), entities conduct activities. On the lowest level, the outputs of these activities should be measured, as shown in the right-hand column. Critical questions during this process, include:

- a) what kind of activities were conducted?

ANALYST, COMMANDER, AND STAFF INTERVIEWS

- b) what was the intent of these activities? and
- c) were there any difficulties encountered in executing these activities? More generally, information is gathered on how the planned activities were executed and what the output of these activities. Based on these measurements, missions or the execution of the activities can be adapted.

On the middle level of the Figure below, the outcomes of the plans are measured. At this stage, the outcome of the plan is measured against its stated objectives in the short term, by asking a question like: were these short term objectives achieved? In order to answer this question, assessment information collected at activity level is used next to the assessment of the short term effects. The results of these outcome assessments are used to adapt and revise the OPLAN in the next iteration. Measurement of mission progress is conducted at the highest level in the figure. Mission progress is described as progress on the themes that are described in the long term objectives of the mission design. This measurement is conducted less frequently than the measurement of effects. Besides separate mission progress assessments, the outcome assessments and the output assessments are also input for the measurement of mission progress. The comprehensive mission design is iteratively adapted based on changes in the environment and the measurement of mission progress and the advice that is derived from this measurement.

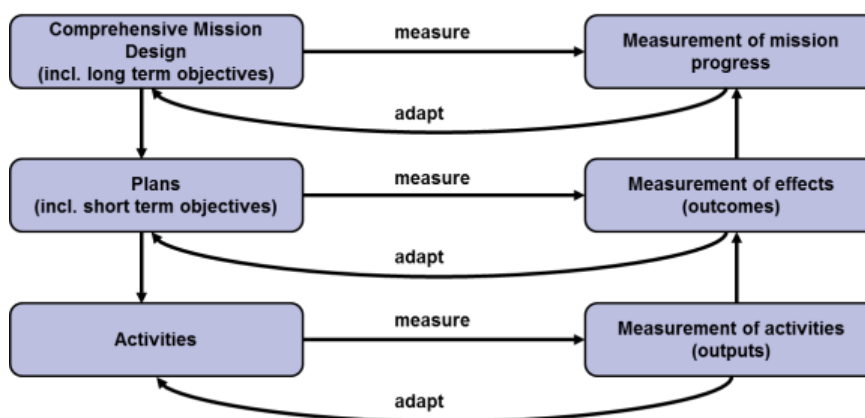


Figure 3-2: Connecting Assessments and Planning.

Such an integrated plan would help to resolve the apparent issues caused by replacements and transfers of authority as well as the eventual transition of security and governance responsibility back to indigenous governments. In Afghanistan, participants repeatedly noted that short rotations and changes in command have the disadvantage of limiting the efficacy of assessors since, as a US Lieutenant Colonel noted, personnel tend to spend the first three months of a nine-month rotation acclimating, the next three months operating at 110%, and the last three months preparing to return home. Very little time is spent learning from the prior group of soldiers in an area of operation and the absence of an assessments package increases the likelihood that the incoming commander will ‘reinvent the wheel.’ As a consequence, a trainable, complete assessments framework that plugs into the campaign planning process, incorporates findings from other government agencies, and allows for successful transitions between military units and from the military to the host government would greatly improve the assessments process. Indeed, from an assessment point of view, a well-managed relationship with the host country’s civilian authorities and armed forces could contribute positively to the process of assessing progress and adding to trustworthy results. In order to “take advantage of this (local) expertise, there is a need to develop tools and methods which are appropriate for the specific context and level of education of the partners [22].”

Strong leadership and ownership of assessments at a senior-level will improve the final product. One American interviewee noted that, in Iraq “it was not difficult to put together [an assessment plan] because of the strong leadership and intellect of the leader, General Petraeus.” Likewise, by soliciting feedback and inputs from the military’s civilian partners, assessments will be more thorough and complete, with a broader focus and a more holistic perspective on the operations theatre.

3.5 CONCLUSIONS

Overall, the development of a consistent assessment framework, sufficient training in that framework, and the appropriate resourcing of military personnel and their civilian counterparts to properly collect the required information are vital to improving current production and use of comprehensive assessments in multi-national NATO operations. None of the recommendations drawn from interviews and workshops are meant to belittle the current practice of assessments in multi-national operations. In fact, the assessments that have been conducted and the information that has been collected, despite less than favourable operational environments, are noteworthy. Clearly comprehensive assessments conducted in conflict areas will remain challenging and require creative and insightful application of solid research methods.

Interview results suggest the need for further study and field tests to explore systematic and defensible methods for evaluating effects in conflict environments *that are appropriate to specific conditions in theatre and available personnel*. This is an issue relevant to each of the countries represented in this study, and becomes even more important in the context of whole of government and multi-national operations in conflict zones. One area of fruitful research would be to establish a core list of indicators applicable to multi-national operations environments. It is important to identify the causal relationships underlying assessment findings, particularly the mechanisms that led to a program’s success or failure: “otherwise, it is difficult to discover credible shortcomings and address necessary changes in either the design or the execution of the plan” [22] or to replicate successful operations in new locales. One Canadian interviewee recommended the use of the “Commander’s Social Science Advisory Team” or “a team designed to coordinate the collection of information as a group of trained and qualified personnel, including placement of “specialised” troops who collect information under an advisory type control. This initiative is not intended to stop the collection of information by troops on the ground, only to coordinate and properly vet it.”

In multi-national operations environments with different deployment configurations than those discussed in Afghanistan, attention should be paid to link different echelons of command in a systematic and consistent fashion so that the sense and value of the assessment remains the same at all levels. In Afghanistan, the Swedes noted that there was poor coordination concerning assessment results between different PRTs and RC North. In order to rectify this situation, higher command levels should consider adopting a more systematic assessment approach that includes consistent collection and data types, while allowing for flexibility and tailoring at lower echelons based upon the commander’s assessment of the situation. There is good reason to suspect that this would ease data collection and analysis demands on lower echelons while improving the precision of assessment products. As one Canadian interviewee noted, “data collection, fusion, and storage need to be better coordinated and managed. A centralised fusion and data management cell would be ideal; however, at a minimum, there needs to be a generally available “index” describing what data is available in which stovepipe.”⁹ Standardised definitions of metrics (i.e., what precisely constitutes an IED find? That is, does a discovery of a cache of unemplaced IED components count?) are also critical to ensure maximum utility of the data being collected.”

Finally, the results presented here suggest that adoption of a common lexicon for assessments that is mindful of the gulf between civilian and military planners, as well as language barriers between different militaries and the indigenous population; common definitions and benchmarks minimise confusion.

⁹ Since the time of these interviews a system has been developed to identify available datasets. More information is available at http://www.datacards.org/mediawiki/index.php/Main_Page



Chapter 4 – ASSESSMENTS BY NON-MILITARY ACTORS

4.1 INTRODUCTION

The goal of Chapters 4 and 5 are to show how readily available data sets can be used to determine a small set of core indicators that can be used to show macro change in a geographical area over time. Chapter 4 focuses on showing that there are data sources from international, non-governmental and government organisations that assess mission areas, country development or societal change already in existence. By using a sample of convenience, we provide a short list and some background on these readily available data sets. The sample of sets demonstrates the availability of consistently measured and reliable indicators across many areas of focus including human security, stability, improving standards of living, increased health, and basic human rights. Below, we present some of the examples provided by different organisations. Then, using one of the data sets, Chapter 5 details the method for variable selection and the statistical analyses used to validate the core set (i.e., derive good indicators). It is important to note that there is no universal set, but that this one is a viable and valid one. And, there are vast data sets that already exist that can be used.

4.2 WORLD HEALTH ORGAISATION (WHO)

The aim of the World Health Organisation is described on the WHO website¹:

“WHO is the directing and coordinating authority for health within the United Nations system. It is responsible for providing leadership on global health matters, shaping the health research agenda, setting norms and standards, articulating evidence-based policy options, providing technical support to countries and monitoring and assessing health trends.”

4.2.1 WHO Statistical Information System (WHOSIS)

WHOSIS, the WHO Statistical Information System, is an interactive database bringing together core health statistics for the 193 WHO Member States. It keeps track of more than 70 national core indicators of health (e.g., mortality, morbidity, risk factors, health systems). Detailed information can be found about each measure, how it is measured, informational sources and estimation methods (the so-called indicator compendium). The following 6 broad categories are used:

- Demographic and socioeconomic statistics;
- Health service coverage;
- Health systems resources;
- Inequities in health care and health outcome;
- Mortality and burden of disease; and
- Risk factors.

Below, an example is given using the latest available data on a sub-selection of these indicators for Congo and the Netherland. Table 4-1 gives a representation of these data.

¹ <http://www.who.int/research/en/>

Table 4-1: Examples of Relevant Measures for Standard of Life, Comparing the Latest Available Data (November 12, 2009) for Congo and the Netherlands, found in WHOSIS.

Indicator	Congo	Year	Netherlands	Year
Adult literacy rate (%)	84.4	2004		
Net primary school enrolment ratio female (%)	52.0	2006	97.0	2006
Net primary school enrolment ratio male (%)	58.0	2006	99.0	2006
Antenatal care coverage - at least four visits (%)	75.0	2005		
Births attended by skilled health personnel (%)	83.0	2005	100.0	2006
Community and traditional health workers density (per 10 000 population)	<1.0	2004		
General government expenditure on health as percentage of total government expenditure	4.0	2006	16.4	2006
Hospital beds (per 10 000 population)	16.0	2005	50.0	2003
Number of nursing and midwifery personnel	3672.0	2004	239172.0	2006
Number of physicians	756.0	2004	60519.0	2005
Nursing and midwifery personnel density (per 10 000 population)	10.0	2004	146.0	2006
Physicians density (per 10 000 population)	2.0	2004	37.0	2005

4.2.2 Global Health Atlas²

In a single electronic platform, the WHO's Communicable Disease Global Atlas is bringing together standardised data and statistics (e.g., for infectious diseases) at country, regional, and global levels. The analysis and interpretation of data are further supported through information on demography, socioeconomic conditions, and environmental factors. There is an interactive map-option³ that allows users to select geographic areas of interest and create maps of diseases, the location of health facilities, schools, roads and geographic features. In addition, there is access to static maps providing some regional information. An example of a static map is provided in Figure 4-2.

Other availability and displays of health related information at the website include:

- The prevalence of chronic diseases which is a database displaying information according to chronic diseases and their risk factors, such as alcohol consumption, and type of diet, etc.;
- Regional offices, provides country health fact sheets with information about a country as compared to the region it is part of. Information includes: life expectancies, causes of death, HIV-prevalence, access to improved sanitation, and the Millennium Development Goals; and
- Global Burden of Disease (GBD) provides a comprehensive and comparable assessment of mortality and loss of health due to diseases, injuries and risk factors for all regions of the world. The overall burden of disease is assessed using the Disability-Adjusted Life Year (DALY), a time-based measure that combines years of life lost due to premature mortality and years of life lost due to time lived in states of less than full health. The WHO GBD project draws on a wide range of data sources to quantify global and regional effects of diseases, injuries and risk factors on population health. The latest assessment of GBD for 2004 is available⁴, together with updated projections to the year 2030, and documentation of methods and data sources. Manuals, software and other resources are also provided for carrying out disease studies.

² <http://apps.who.int/globalatlas/default.asp>.

³ <http://apps.who.int/globalatlas/interactiveMapping/MainFrame2.asp>.

⁴ http://www.who.int/entity/healthinfo/global_burden_disease/2004_report_update/en/index.html.

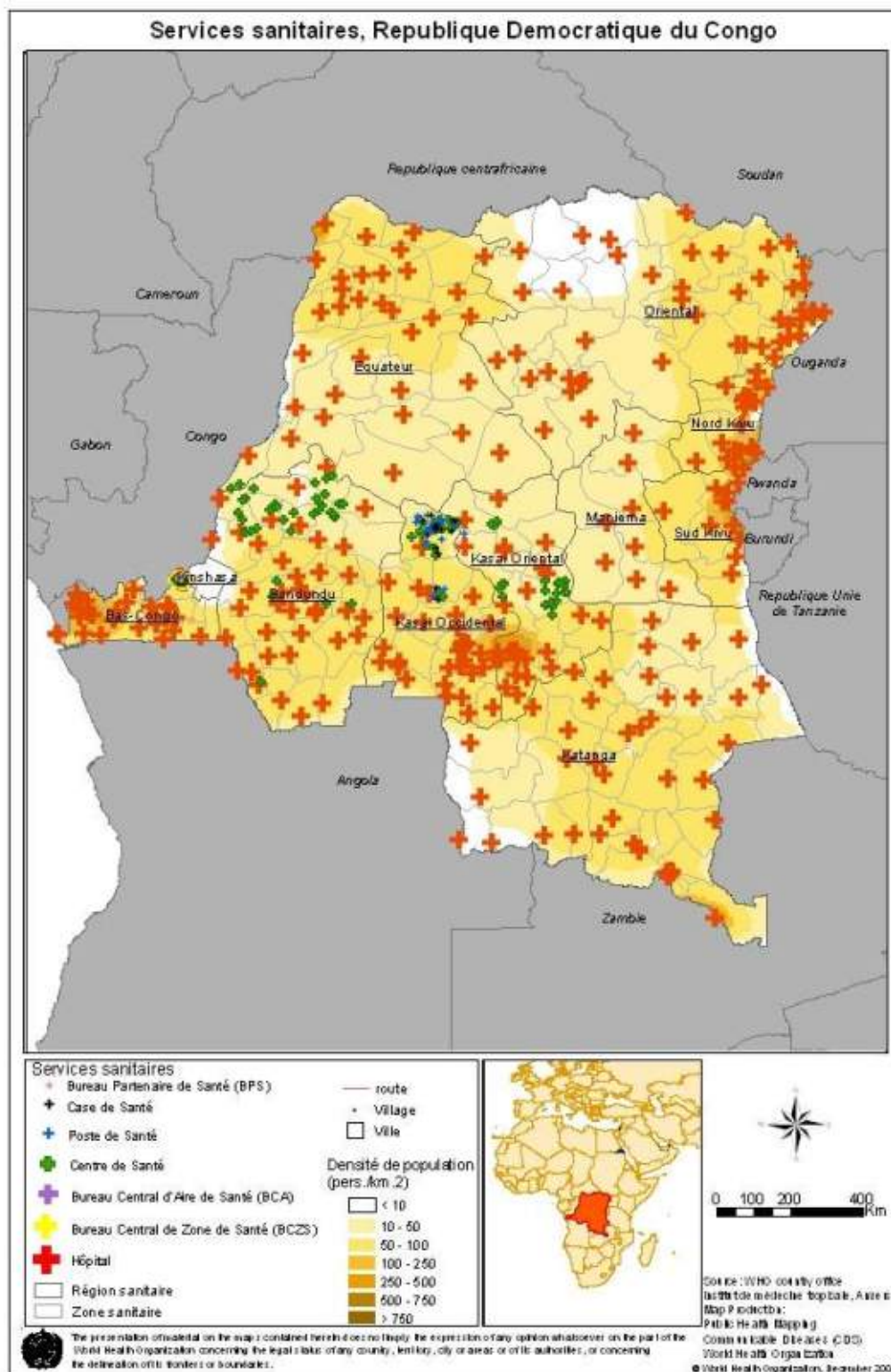


Figure 4-2: Example of a Static Map.

http://gamapserver.who.int/mapLibrary/Files/Maps/CD_overview.jpg

4.3 FREEDOM HOUSE (FH)

The Freedom House⁵ website says that:

“Freedom House supports democratic change, monitors freedom, and advocates for democracy and human rights around the world. We support nonviolent civic initiatives in societies where freedom is denied or under threat and we stand in opposition to ideas and forces that challenge the right of all people to be free. Freedom House amplifies the voices of those fighting for freedom in repressive societies. We work directly with democracy and human rights advocates in their own countries and regions. These reformers include human rights defenders, civil society leaders and members of the media. Freedom House’s programs provide these advocates with resources that include training, expert advice, grants and exchange opportunities. We press the United States, other governments, international institutions and regional bodies to adopt consistent policies that advance human rights and democracy around the world.

The foundation of Freedom House’s work is its analysis. We evaluate the components of freedom and leverage our analytical work to strengthen our advocacy and action efforts. Freedom House’s rigorous research methodology has earned the organisation a reputation as the leading source of information on the state of freedom worldwide.”

4.3.1 Databases

The *Freedom in the World* survey is an annual evaluation of the state of global freedom. It measures freedom according to two categories – political rights and civil liberties. Political rights are about letting people participate freely in the political process. Civil liberties are about the freedoms of expression and belief, associational and organisational rights, rule of law, and personal autonomy without interference from the state.

The survey consists of analytical reports and numerical ratings for 193 countries and 14 select territories. Each country and territory report includes a section summarising the current state of political rights and civil liberties. The political rights questions are grouped into three subcategories: Electoral Process, Political Pluralism and Participation, and Functioning of Government. The civil liberties questions are grouped into four subcategories: Freedom of Expression and Belief, Associational and Organisational Rights, Rule of Law, and Personal Autonomy and Individual Rights.

Data can be accessed by Tables comparing countries, and in country reports.⁶

4.4 UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANISATION (UNESCO)

UNESCO⁷ “works to create the conditions for dialogue among civilisations, cultures and peoples, based upon respect for commonly shared values. It is through this dialogue that the world can achieve global visions of sustainable development encompassing observance of human rights, mutual respect and the alleviation of poverty, all of which are at the heart of UNESCO’s mission and activities” (cited from website).

⁵ <http://www.freedomhouse.org/>.

⁶ <http://www.freedomhouse.org/template.cfm?page=21&year=2008>.

⁷ <http://www.unesco.org>.

4.4.1 Institute for Statistics

“The UNESCO Institute for Statistics (UIS) is the statistical branch of the United Nations Educational, Scientific and Cultural Organisation (UNESCO). The Institute produces the data and methodologies to monitor trends at national and international levels. It delivers comparative data for countries at all stages of development to provide a global perspective on education, science and technology, culture, and communication” (cited from website).

The Data Centre⁸ contains over 1,000 types of indicators and raw data on education, literacy, science and technology, culture and communication. The UNESCO Institute for Statistics (UIS) collects the data for more than 200 countries. Often, sample-technology is used to obtain data representative on country level. Data can be downloaded in Excel files.

4.5 WORLD BANK

The World Bank⁹ (WB) is a partnership between 187 countries to reduce poverty and support development. They offer support to developing countries through policy, advise, research and analysis, and technical assistance. For example, their open data website offers free access to comprehensive, downloadable indicators about development in countries around the globe. The data record gives access to over 7000 indicators. They can be viewed by country, topic, or indicators.

⁸ <http://stats.uis.unesco.org/unesco/tableviewer/document.aspx?>

⁹ www.worldbank.org.



Chapter 5 – CORE VARIABLES TO ASSESS CHANGE OVER TIME

As was demonstrated in Chapter 4, there are many organizations using hundreds of indicators to determine a global index of a nation's "health or wellness". The purpose of using such indicators is to assess change in a chosen geographical area over time. Currently, there is a very large number of different indicators used with different metrics and methods of collection and analyses, which makes it difficult to apply them to make timely and trustworthy measurements and assessments of outcomes in multi-national operations. The Task Group's aim here is to show that a very small set of core indicators can make a global appraisal about the condition of the area with enough precision to provide a valid and reliable assessment.

In this Chapter, we present a method for determining a small set of core variables and demonstrate (with readily available data) how these variables can be used to show macro change in a geographical area for a specific period of time. The importance of having a standardized set has been raised in previous chapters; however, it is worth noting that if the core variables are collected for each NATO mission, then we can readily make comparisons across NATO missions, something that currently is not possible. It is important to note that long-term strategic change that we are aiming to assess cannot be observed due to a specific and recent tactical military intervention - that type of change requires local measurement. However, the two levels of assessment (strategic and tactical) should be complimentary. That is, by conducting trend analysis with the core set, certain large-scale military interventions should be identifiable.

5.1 APPROACH

From the selected databases discussed in Chapter 4, all variables from each database were listed. The list resulted in thousands of items. Many of the data sets included similar variables (e.g., birth rate; death rate), therefore duplicate variables were removed. This reduced the original set to a sub-set of 3,317 variables. Following this, each variable was categorized under the following high level constructs that are pertinent to NATO operations: Economy, Education, Freedom, Health, Infrastructure, Law, Security, and Social Well-being. It is important to note that these categories are not mutually exclusive. That is to say, the variables used as indicators for a particular construct in one data set may be used for a different construct in another.

Intensive group discussion was used to scrutinise the list; careful attention was paid to the pre-determined criteria discussed in Chapter 2 (S.M.A.R.T.) and also based on the following additional attributes: the item must be observable, culture-free, counted the same way by everyone, and be independent (i.e., not a composite variable or an index). The discussion resulted in a condensed subset of 119 variables. The expert team then individually ranked from most important to least important what they believed to be the top 30 variables out of the subset of 119. This new set was once again reduced by eliminating duplications leaving a reduced set of 53 variables. In a second down-selection of this sub-set, variables were eliminated that were only chosen by one expert. The final list consisted of 33 variables and is presented in Table 4-1. It is important to note that no indicators of *freedom* are represented in the final list. This is due to the fact that indicators for this construct are not measured objectively, and are culturally biased. Future work must examine metrics that can objectively quantify this construct.

5.2 THE SET OF INDICATORS

The World Bank data sets were chosen for the demonstration as they had the most exhaustive set that matched the core set identified by the SMEs; other databases were not used in combination with the World Bank set as it was extremely difficult and sometimes impossible to merge the sets (i.e., convert the data to a common metric). Relevant operationalizations of the variables from the World Bank data that matched the final variable list are shown in Table 5-1.

CORE VARIABLES TO ASSESS CHANGE OVER TIME

Twelve of the items could not be represented by data from the World Bank database, leaving 21 variables in total. Variables marked with an asterisk indicate additional constraints on the proxy measure; for example, the variable called ‘improved sanitation facilities’ has not necessarily been operationalized to mean the same thing as having a private toilet. Unfortunately, specific information on the operationalizations of the variables and the data collection methods are not provided in sufficient detail at the WB website. This is problematic as it is possible that the specified criteria for inclusion into the list of core variables may not have been met. Nevertheless, there is enough representation of a potential core set to present a meaningful demonstration as to how a small subset of variables can show global change. Note that this demonstration is intended for an audience that understands statistical procedures applied to social science problems. SPSS® software was used for the analyses.

Table 5-1: Set of Indicators and SPSS Variable Names Based on World Bank Data.¹

#	Operationalised Item	Category	Nation Level Proxy Variable Name (as listed in WB data set)
1	Total population	Demographic	SP_POP_TOTL
2	Number of males	Demographic	SP_POP_TOTL_MA
3	Number of females	Demographic	SP_POP_TOTL_FE
4	Number of households	Demographic	No proxy available
5	Number of shops*	Economy	IC_BUS_TOTL
6	Average price of "product" gasoline	Economy	EP_PMP_SGAS_CD
7	Number of markets	Economy	No proxy available
8	Labor participation rate, female (% of female pop aged 15+)	Economy	SL_EMP_TOTL_SP_FE_ZS
9	Labor participation rate, male (% of male pop aged 15+)	Economy	SL_EMP_TOTL_SP_MA_ZS
10	Arable land (% of land area)	Economy	AG_LND_ARBL_HA
11	Household income (average)	Economy	NE_CON_PETC_CD
12	Household income (standard deviation)	Economy	No proxy available
13	Male Adult literacy rate	Education	SE_ADT_LITR_MA_ZS
14	Female Adult literacy rate	Education	SE_ADT_LITR_FE_ZS
15	% of households with private toilet*	Health	SH_STA_ACSN
16	% fully immunized (DPT)	Health	SH_IMM_IDPT
17	% of population with access to safe water*	Health	SH_H2O_SAFE_ZS
18	Number of medical doctors	Health	SH_MED_PHYS_ZS
19	Adult mortality rate female (CM)	Health	SP_DYN_AMRT_FE
20	Adult mortality rate male (CM)	Health	SP_DYN_AMRT_MA
21	Infant mortality rate (under 5)	Health	SP_DYN_IMRT_IN
22	Number of hours electricity	Infra structure	No proxy available
23	Roads, paved (% of total roads)	Infra structure	IS_ROD_PAVE_ZS
24	Number of public transportation passengers	Infra structure	No proxy available
25	Number of court cases (criminal & civil)	Law/Security	No proxy available
26	Number of Law Enforcing Officers	Law/Security	No proxy available
27	Number of incarcerated individuals	Law/Security	No proxy available
28	Number non accidental deaths	Law/Security	No proxy available
29	Number of cultural events/public celebrations	Social	No proxy available
30	Number of households with washing machine	Social	No proxy available
31	Number of daily newspapers available in shops	Social	IT_PRT_NEWS_P3
32	Number of available broadcasting stations (radio & tv)	Social	No proxy available
33	Number of internet connections	Social	IT_NET_BBND

¹ Rows highlighted in yellow indicate that no proxy variable was available for the analysis.

5.3 DEMONSTRATION

Consider the following: if one assumes that the nations in the world are regions in an area of operations we can then use national level data to demonstrate the argument that a small set of indicators can be used to assess macro change over time. Following an examination of the univariate descriptive statistics of the data used (see Table 5-2) to ensure the set is normally distributed, we performed a Principal Component Analysis² to test the assumption of orthogonality of the indicators; principal components analysis is appropriate when you have obtained measures on a number of observed variables and wish to develop a smaller number of variables (principal components) that will account for most of the variance in the observed variables. The scope of the dataset is from 1960 - 2009 and consists of one record per country (n = 237 nations) per year and 21 variables per record. In total, 11,850 records were included in the analysis, but for a host of variables missing values were replaced by the means.

Table 5-2: Descriptive Statistics for the Demonstration Dataset.

#	Item	Mean	Std. Deviation(a)	Analysis N(a)	Missing N
1	Population Total	137628938.7728	541948563.49939	11850	367
2	Population Total Males	74890297.7914	273465620.36188	11850	1235
3	Population Total Females	73940614.3960	267142385.91084	11850	1235
5	Total businesses registered (number)	454591.1594	176349.17224	11850	11367
6	Pump price for gasoline (US\$ per litre)	.7475	.12607	11850	10432
8	Employment to population ratio, 15+, female (%)	45.0700	8.30113	11850	8304
9	Employment to population ratio, 15+, male (%)	70.2518	5.47797	11850	8304
10	Arable land (hectares)	6513376.9097	18173952.85646	11850	3223
11	Household final consumption expenditure, etc. (current US\$)	371493997137.1737	1426028602396.62600	11850	4170
13	Literacy rate, adult male (% of males ages 15 and above)	82.1749	3.96442	11850	11296
14	Literacy rate, adult female (% of females ages 15 and above)	72.2281	5.55311	11850	11296
15	Improved sanitation facilities (% of population with access)	68.4986	8.92638	11850	10854
16	Immunization, DPT (% of children ages 12-23 months)	75.0930	16.22944	11850	6402
17	Improved water source (% of population with access)	83.1011	5.52778	11850	10833
18	Physicians (per 1,000 people)	1.5308	.79868	11850	8420
19	Mortality rate, adult, female (per 1,000 female adults)	156.5227	69.87809	11850	7502
20	Mortality rate, adult, male (per 1,000 male adults)	236.5657	70.38249	11850	7501
21	Mortality rate, infant (per 1,000 live births)	41.8415	24.05665	11850	8081
23	Roads, paved (% of total roads)	50.8975	14.52975	11850	9559
31	Daily newspapers (per 1,000 people)	131.2272	32.99887	11850	11154
33	Fixed broadband Internet subscribers	967172.7383	1916289.75365	11850	10177

Note: For each variable, letter 'a' denotes missing values which are replaced with the variable mean.

² Principal components analysis is a procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values that are linearly uncorrelated variables called principal components. The transformation is defined so that the first component has the largest variance (accounts for as much of the variability in the data as possible). Each succeeding component has the next highest variance possible under the constraint that it is uncorrelated with the preceding components.

CORE VARIABLES TO ASSESS CHANGE OVER TIME

In the Principal Components Analysis, 7 orthogonal dimensions based on the 21 variables at our disposal were extracted. In reality, the number of components extracted in a principal component analysis is equal to the number of observed variables being analysed, in this case 21. However, in most cases, only the first few components account for meaningful amount of variance, so only these first few components are retained, interpreted, and used in subsequent analyses, such as in multiple regression analyses. These 7 dimensions, based on the 21 variables, can distinguish nations and years in a 7 dimensional space, allowing for sufficient degrees of freedom to discern differences (see Table 5-3).

Table 5-3: Total Variance Explained for Demonstration Dataset.

Component	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.680	17.525	17.525	3.362	16.010	16.010
2	3.393	16.158	33.683	2.366	11.268	27.278
3	1.848	8.798	42.481	1.966	9.364	36.641
4	1.411	6.718	49.199	1.955	9.310	45.951
5	1.377	6.560	55.758	1.873	8.919	54.870
6	1.307	6.225	61.983	1.421	6.767	61.637
7	1.229	5.854	67.837	1.302	6.200	67.837

Extraction Method: Principal Component Analysis.

To interpret this data at a global level, Table 5-4 shows which variables contribute to the 7 principal components.

Table 5-4: Rotated Component Matrix for Demonstrated Dataset.

#	Rotated Component Matrix(a)	Component						
		1	2	3	4	5	6	7
3	Population Total Females	.985	.013	-.052	.005	-.010	.011	-.045
1	Population, total	.983	.013	-.054	.004	-.012	.014	-.048
2	Population Total Males	.983	.013	-.054	.005	-.010	.012	-.047
11	Household final consumption expenditure, etc. (current US\$)	.662	-.079	.106	.003	.021	.127	.146
20	Mortality rate, adult, male (per 1,000 male adults)	-.030	.933	-.117	-.025	-.065	-.005	-.018
19	Mortality rate, adult, female (per 1,000 female adults)	-.022	.929	-.185	-.049	-.092	-.030	.004
21	Mortality rate, infant (per 1,000 live births)	.006	.604	-.145	-.055	-.352	-.040	-.102
9	Employment to population ratio, 15+, male (%)	.044	-.003	-.805	-.074	-.064	.004	.019
23	Roads, paved (% of total roads)	.013	-.100	.632	-.030	.165	.009	.228
8	Employment to population ratio, 15+, female (%)	.041	.134	-.628	-.012	-.006	.067	.462
16	Immunization, DPT (% of children ages 12-23 months)	.022	-.203	.441	.105	.153	.048	.139
18	Physicians (per 1,000 people)	.008	-.272	.400	.034	-.004	.021	.241
13	Literacy rate, adult male (% of males ages 15 and above)	.012	-.055	.078	.980	.060	.003	-.004
14	Literacy rate, adult female (% of females ages 15 and above)	.002	-.057	.073	.979	.086	.003	.006
17	Improved water source (% of population with access)	.012	-.158	.143	.058	.915	.020	-.001
15	Improved sanitation facilities (% of population with access)	-.020	-.157	.157	.087	.907	.009	-.004
5	Total businesses registered (number)	-.015	.032	.012	.010	.021	.716	-.007
33	Fixed broadband Internet subscribers	.028	-.043	-.001	.017	.043	.698	.137
10	Arable land (hectares)	.113	-.044	.005	-.021	-.034	.627	-.136
6	Pump price for gasoline (US\$ per liter)	.012	.079	.090	.010	.066	-.003	.700
31	Daily newspapers (per 1,000 people)	-.007	-.228	.126	-.012	-.066	-.028	.627

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalisation. a Rotation converged in 5 iterations.

In order to show the changes over time, correlations were calculated between the variable year and each of the variables in the demonstration set (see Table 5-5). It was expected that there would be a significant positive correlation between year and population numbers, which is supported by the analyses. As another example, we find adult literacy rates are negatively correlated with mortality rates for all three population sectors (males, females, and infants). Infant mortality rates are also negatively correlated with numbers of physicians, improved water sources, and improved sanitation facilities. The relationships between these principal components (health and social well-being) reveals the kinds of long term societal changes we expect to see in flourishing societies. This is why military interventions tend to focus on producing changes in these sorts of variables. For example, development strategies in Afghanistan have included literacy programs, immunization programs, the building of wells and the paving roads. Short term tactical assessments, such as counting the number of wells drilled every 6 weeks, cannot show the long term benefits of such strategies.

CORE VARIABLES TO ASSESS CHANGE OVER TIME

Table 5-5: Correlations between Year and Other Variables.

#	Item	N	Pearson r
1	Population Total	11483	0.056**
2	Population Total Males	10615	0.061**
3	Population Total Females	10615	0.060**
5	Total businesses registered (number)	483	0.048
6	Pump price for gasoline (US\$ per litre)	1418	0.411**
8	Employment to population ratio, 15+, female (%)	3546	0.075**
9	Employment to population ratio, 15+, male (%)	3546	-0.064**
10	Arable land (hectares)	8627	0.016
11	Household final consumption expenditure, etc. (current US\$)	7680	0.127**
13	Literacy rate, adult male (% of males ages 15 and above)	554	0.268**
14	Literacy rate, adult female (% of females ages 15 and above)	554	0.262**
15	Improved sanitation facilities (% of population with access)	996	0.078*
16	Immunization, DPT (% of children ages 12-23 months)	5448	0.431**
17	Improved water source (% of population with access)	1017	0.139**
18	Physicians (per 1,000 people)	3430	0.310**
19	Mortality rate, adult, female (per 1,000 female adults)	N	-0.008
20	Mortality rate, adult, male (per 1,000 male adults)	11483	-0.080**
21	Mortality rate, infant (per 1,000 live births)	10615	-0.280**
23	Roads, paved (% of total roads)	10615	0.154
31	Daily newspapers (per 1,000 people)	483	0.073
33	Fixed broadband Internet subscribers	1418	0.132**

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

By plotting the data over time (see Figure 5-1) we can examine the percentage of males and/or females employed between the years 1991 and 2008. Globally we see a steady increase in the female employment rate and a slight decline in the male employment rate (based on the current dataset, in which the data are values per year and per nation). This steady increasing trend of female employment rates shows the macro change of this item over time. A true representation of this variable cannot be captured by measuring fluctuating employment rates from short term impact projects four times per annum in a village in Afghanistan, in an effort to show that the labor participation rates are changing as a result of military intervention.

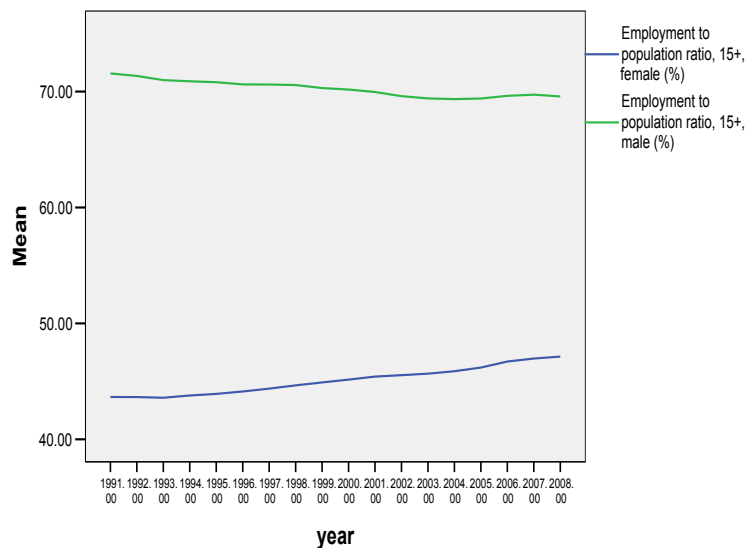


Figure 5-1: Male and Female Employment Rates as a Function of Time.

5.4 CONCLUSIONS

This demonstration shows that a small set of core indicators could be very helpful in reflecting societal change over time. We identified 33 possible indicators, based on 8 strategic level categories common to comprehensive approach interventions, although we were only able to identify non-subjective measures for 7. Using data available for 21 of 33 variables, we demonstrated that they can be used to show the current system-state of a geographical area. The demonstration was partial, however, due to the inability to find usable data for all 33 variables identified (there were no suitable measures identified for the construct of *freedom*). Also, it was not possible to determine the appropriate spatial and temporal resolution required for collection purposes at this time.

It is important to understand that these variables may not always be directly linked to assessments required by a commander's specific mission (tactical level campaign assessments), but they should detect changes in higher level and longer term goals. In fact, the activities conducted and effects achieved on missions of 3 – 12 months in length will most certainly produce 2nd, 3rd or 4th order effects, which most likely will only be detectable on longer timescales. The current practice of continual revisions of MOEs and their measures to identify immediate change (i.e., are we winning) as a result of specific interventions will never allow for the identification of slow nation building progression. Essentially, commanders should focus on MOPs, as this is truly what they are in a position to assess. Additionally, the troops in the field, even when trained, are not always a viable data collection source. Ideally, MOEs should be left to other agencies that are able to examine them systematically and for the long term.

In addition to revealing the long-term effects of military interventions, a set of core variables will also allow for comparisons between missions. Currently, evaluations between missions are extremely difficult (due to different measures, metrics, and methods) if at all possible to conduct. Understanding efforts at a global level will help determine where resources and efforts should be allocated.

Further work is also required to determine what other measures currently in use for specific campaign assessment could be analysed to establish which items have the greatest predictive validity. For example, if there are dozens of different indicators for the construct of safety, it should be determined which one or two indicators are the most indicative of that construct (i.e., provide the greatest predictive and explanatory power) to reduce time and effort expended for data collection and assessment efforts.

CORE VARIABLES TO ASSESS CHANGE OVER TIME

In summary, a core set of indicators would be very beneficial to compliment current tactical level campaign assessment. Further research is required to determine:

- a) the complete comprehensive set that provides the greatest explanatory and predictive power,
- b) over what geographical and temporal resolution Core Indicators should be measured. A final set, once selected and validated, must be integrated in NATO's COPD [15] and Assessment Handbook [16].

Chapter 6 – CONCLUSIONS AND RECOMMENDATIONS

6.1 INTRODUCTION

Through interviews, questionnaires, and workshops with the consumers and producers of campaign assessments in multi-national operations and through a demonstration of available data sources from international organisations, the HFM-185 Task Group illustrates the need for continued research to improve the assessment of multi-national missions.

Chapter 2 briefly described what constitutes good assessment and measurement (work that has been much more thoroughly investigated through other groups, panels and committees), while Chapter 3 documented the inconsistent use of good assessment and measurement practices in theatres of operations. Chapter 4, in turn, provided a brief synopsis of data sources and used that data to show in Chapter 5 the potential utility of a having a core set of metrics to compliment ongoing campaign assessments. Overall, the current state of multi-national assessments will be improved through the development and institution of the following enablers:

- Standard assessment procedures and frameworks for use throughout NATO/ISAF and across different multi-national operations;
- A common lexicon and set of definitions across NATO/ISAF member nations provided through the relevant NATO documentations COPD [15] and Assessment Handbook [16];
- A core set of metrics to supplement ongoing campaign MOPs and MOEs included as an Annex to the Assessment Handbook [16];
- Integrated assessment and campaign planning, while also improving the coordination between civilian and military assessment activities within multi-national operations;
- The consistent training of relevant assessment personnel in the standardised assessment approach or framework during command and staff training and before deployment;
- Accessible, integrated information systems across NATO/ISAF member nations that provides information access at the lowest possible level of clearance; and
- Senior leaders who take ownership of the assessment process and instantiate a culture of objective campaign assessments.

The following pages provide a more complete characterisation and discussion of each of these recommended enablers. The implementation of some of these will likely be more difficult than others. Based on empirical data and the Task Group members' own experiences, these enablers are likely to address some of the main challenges identified throughout the report, while improving the assessment process and ensuring results that are objective in order to improve the iterative cycle of planning and assessment.

6.2 REPLICABLE ASSESSMENT STANDARDS AND/OR DATA FRAMEWORK

The research findings highlighted an absence of NATO assessment practices that allowed data to be collected throughout a mission in a consistent and replicable manner. Overarching organisational structures and assessment frameworks are needed to direct the way in which assessments are produced and used inside the organisations, and operationalize the mission plan into metrics and give specifics on data collection and frequency.

CONCLUSIONS AND RECOMMENDATIONS

In terms of organisational factors, the findings showed that mission plan objectives and guidance on assessments were not consistently disseminated to lower echelons. Specifically, lower level commanders developed assessment frameworks in an inconsistent manner and often prioritised their own tactical concerns over the strategic aspects of the assessment plan.

With regard to framework factors, the findings showed that assessment frameworks were inconsistently produced across missions and commands, meaning that assessment taskings were often ad hoc in nature with varying timelines, had different underlying objectives, and were inconsistent or lacking in the extent of operationalization into metrics. Furthermore, there was a lack of reliable benchmarks and baselines against which assessments could be compared over time. Additionally, as a result of changes in command and other factors, metrics changed so frequently that trend analysis beyond battle damage assessments (e.g., counting IED strikes and finds) was almost impossible.

In order to remedy these concerns, the study has identified two main recommendations:

- 1) development of standardised assessment guidance in policy or doctrine documents across different nations, organisations, and commands, and
- 2) development of a standardised assessment framework from the outset of operations. This would improve standardisation of assessment practice in missions, thus improving the quality, legitimacy, and reliability of assessments produced.

6.3 COMMON LEXICON AND/OR TERMINOLOGY

Definitions and wordings are often understood differently within the assessment community in multi-national interventions. Definitions may have multiple meanings depending on who is using them and the context in which they are used. A key finding in this study is that there exists no commonly used lexicon and understood terminology regarding assessment. Consequently, assessment data and outcomes often get “lost in translation” across units of different nationalities and between military and civilian actors.

The Task Group recommend that organisations acting in multi-national operations support the use of a lexicon helping the assessment community to get better knowledge about the terms used by the different organisations. For instance, the creation of comparison tables showing the different lexicons used by organisations taking part in an operation can contribute to increased understanding of various cultures and practices. Such knowledge will most likely improve the ability to cooperate. In this regard it is important to show where there are differences from the Development Assistance Committee (DAC) of the Organisation for Economic Cooperation and Development (OECD) international standards.

It is imperative that terms and definitions used in specific missions always are a part of introductory briefings and training prior or in the beginning of service in theatre. These lexicons should also be easily located in common reference sources (i.e., NATO Assessments Handbook and NATO COPD). In addition to training and education on assessment and planning, the use of standard operating procedures (or other instructions), which are shared across organisations, can facilitate cooperation and increase efficiency. These will contribute to a common understanding of important planning elements, expressing the different levels of objectives and desired effects used by both planners and assessors within each. Common understanding between planners and assessors is essential for valid and reliable assessment and avoidance of misunderstandings.

6.4 A SET OF CORE VARIABLES

The Task Group determined a preliminary set of 33 core variables. Using data available for 21 of 33 core variables, we demonstrated that a small set of variables could detect societal change over time. In fact, recommendations for such a set of variables have come out of our results (see Chapter 3) as well as numerous workshops and task groups working on assessing both the Iraq and Afghanistan operations.

We identified 33 possible indicators, based on 8 strategic level categories common to comprehensive approach interventions. Using data available for 21 of 33 variables, we demonstrated that they can be used to show the current system-state of a geographical region.

It is important for commanders to understand the importance of data collection efforts for a strategic set of indicators that are complimentary to those that assess specific military actions. The activities conducted and effects achieved on short-term efforts within a larger mission will most certainly produce 2nd, 3rd or 4th order effects, which may only be demonstrable on longer timescales. The continual revision of MOEs and changes in plans due to political requests, individual personalities, and rotation of tasks forces into theatres of operations will never allow for the detection of slow nation building progression. Therefore it is important to ensure that strong leadership of the assessment process understands and communicates the rationale for collecting such measures.

Currently, comparisons between missions are extremely difficult (due to different measures, metrics, and methods) if at all possible to conduct. A set of core variables will also allow for comparisons between missions. Understanding efforts at a global level will help determine where resources and efforts should be allocated or reallocated.

The current demonstration was deficient. Further work is required determine

- a) the appropriate spatial and temporal resolution required and
- b) the comprehensive set that provides the greatest explanatory and predictive power.

In summary, a core set of indicators would be very beneficial to compliment current tactical level campaign assessment. We strongly recommend continuing this research in order to identify a comprehensive set of core variables and validate that set in different operational contexts (i.e., combat missions and humanitarian relief efforts). A final set, once selected and validated, must be integrated in NATO's COPD [15] and Assessment Handbook [16].

6.5 INTEGRATING ASSESSMENTS AND PLANNING

The research discussed in Chapter 3 showed a lack of integration between assessments and planning. Despite the challenge of collecting sufficient data for good assessments, planning rhythms were not in line with assessment rhythms. Assessors and assessment procedures were not known to planners and sometimes assessors were missing altogether. There were also examples of poor communication, coordination, and collaboration between planners and assessors, ending up with assessors working on completely different themes, timelines, and scopes than planners. This made it impossible to provide planners with advice based upon assessments. Even if the different processes were in line, assessment results were often not usable for the planning process due to lack of common understanding of elements in the plan (goals, objectives, etc.).

Assessment needs to become an integral structural part of planning at all levels (from strategic to tactical) and the principles should be incorporated into doctrine. Planning and assessment staffs should work closely together in order to optimise the usage of both processes. However, there is a risk that assessors might

CONCLUSIONS AND RECOMMENDATIONS

become biased towards short term interests (and successes). In other words, assessors may be driven, guided and influenced to assess the priorities of the current operational commander rather than the longer-term strategic goals of the mission. As a consequence of the relatively short deployment duration, it is often impossible to independently notice/observe long-term developments. Therefore, the assessment process should aim for a long-term view and include analysis work with core variables, their metrics and the ways they are measured that are never altered. In this context it is imperative that the different levels of command are integrated with an overarching assessment framework.

Assessments should be related to plans. To ensure common aims, objectives, intents, and risk mitigation operational plans are linked together from the strategic level down to the tactical level. Therefore, the assessment process must be integrated with the planning process on all levels of command. Because the COPD [15] remains the overarching hinge between operational assessment and planning in NATO, it must integrate and detail the link between assessment and planning.

6.6 CONSISTENT TRAINING IN ASSESSMENT

Most participants in the interviews and workshops noted inconsistency in the use and production of assessments between nations, regions, command levels and unit/formation rotations. Contributing factors were no or limited prior experience of assessment, the lack of a common lexicon and terminology, limited understanding of the purpose, implementation and exploitation of assessments, a lack of social science subject-matter expertise and limited understanding of the campaign aims at lower levels. Training in a number of these areas would contribute to a substantial improvement in the quality of assessments.

Those staff who will be involved in assessment should undergo formal training in the purpose of assessment, good practices in its implementation and execution and the specifics of the framework to be used and any tools or software which support it. Specific courses and a further integration of assessment issues into command and staff training courses should be aspired to.

All troops/staff deployed on a multi-national operation should be briefed on the campaign aims, why assessment is important, how it will be achieved and their role in the process.

6.7 INFORMATION TECHNOLOGY AND DATA MANAGEMENT

Respondents pointed to the inadequacies of existing IT infrastructure in theatre, which prevented easy access and search of archive and operational data, and to IT policies, which should mandate archival and classification procedures. In addition, different systems in use across coalition members, members of the military/assessment cells in theatre, and those serving in reachback capacities in their home country, contributed to challenges of data access.

Given the importance of data access to assessment, improved and integrated data and knowledge management systems are required that facilitate information sharing across participating militaries and civilian departments at the lowest level of clearance allowable. In addition, increasing the overlap between incoming and outgoing assessment staffs and commanders would improve the transition of assessment data standards at the local level across deployment transitions by providing for greater awareness of existing data and information sources for the incoming personnel.

6.8 STRONG LEADERSHIP OF THE ASSESSMENT PROCESS

Empirical data demonstrated the need for a single voice to exert strong influence and control over the assessment process. If a senior commander, official, or group takes ownership of the assessment it has the following potential benefits:

- That senior person should have the best understanding of the campaign aims and objectives and thus the assessment is more likely to measure the right things;
- The assessment will be aligned with the understanding of the senior decision maker(s);
- Resources are more easily allocated to the assessment process with the authority of a senior officer or official;
- Changes will not be made arbitrarily to the metrics without approval; and
- Assessment will be more readily integrated into the planning cycle.

Conversely, if the ownership of the assessment is not taken by senior leadership, then it is likely to be viewed as less important by the lower echelons and the process will have less authority. Therefore, assessment should be owned by and coordinated at the level at which the decisions it will be informing will be taken.

6.9 RECOMMENDATIONS

6.9.1 Measurement and Assessment Process Recommendations

NATO and national directives should be updated to give specific guidance about institutionalising assessment procedures in operations in the following areas:

- Developing a strategic assessment plan and transmitting it to subordinate commands;
- Standardised metrics that are collected throughout the duration of a mission by standardised methods;
- Clear and consistent operationalizations of MOEs and MOPs;
- A process to deal with changes in the plan or operational focus;
- A process to deal with changes in command; and,
- While complete standardisation is not desirable, NATO should make some attempt to describe and then harmonise the various approaches across NATO nations.

6.9.2 Operational Practice Recommendations

- Develop an Assessment protocol from the outset of the Mission (Campaign) design;
- Agree with collaborating organisations on an Assessment design and framework; and
- The NATO Operations Assessment Handbook [16] and COPD [15] define many of the terms used in assessment and provide the basis for a common understanding. They should be used as a common basis for national policy, doctrine, and training for assessment.

6.9.3 Future Research Recommendations

- It is strongly recommended that a follow-on Research Task Group consisting of experienced social science and human factors analysts continue research to identify and validate a small set of core variables; and
- Make use of that same follow-on Task Group as an essential research support capability to further develop and document the NATO planning and assessment process.

CONCLUSIONS AND RECOMMENDATIONS



Chapter 7 – REFERENCES

- [1] Diehl, J. G., & Sloan, C. E. (2005). Battle Damage Assessment: The Ground Truth. *Joint Forces Quarterly*, 37, 59-64.
- [2] Effects-Based Approach to Operations. Systems Analysis and Studies Panel (SAS) 070. November, 2007.
- [3] United States Institute of Peace; United States Army Corps of Engineers; and United States Army Peace Keeping and Stability Operations Institute. (2008). *Measuring Progress in Conflict Environments (MPICE): A Metrics Framework for Assessing Conflict Transformation and Stabilisation*. August 2008, Version 1.0.
- [4] Becker, D. C. & Grossman-Vermaas, R. (2011). Metrics for the Haiti stabilization initiative. *Prism*, 2, 145-158. <http://www.ndu.edu/press/metrics-haiti-stabilization-initiative.html>
- [5] Astorino-Courtois, A., & Pagano, S. J. (2009). *Measuring Progress in Conflict Environments (MPICE): Framework Evaluation Report*. Prepared by NSI, INC.
- [6] Organisation for Economic Co-operation and Development. (2008). *Guidance on Evaluating Conflict and Prevention and Peace Building Activities: Working Draft for Application Period*. A joint project of the Development Assistance Committee (DAC) Network on Conflict, Peace, and Development Co-operation and the DAC Network on Development Evaluation.
- [7] Moroney, J. D. P., Marquis, J. P., Quantic Thurston, C., & Treverton, F. (2009). *A Framework to Assess Programs for Building Partnerships*. RAND Monograph. Prepared for the Office of the Secretary of Defense.
- [8] Connable, B. (2010). *An Alternative to Effects-based Campaign Assessment in Afghanistan*. Project Memorandum PM-3618-MCIA. Prepared for the Marine Corps Intelligence Agency by RAND National Defence Research Institute.
- [9] Killcullen, D. J. (2009). *Measuring Progress in Afghanistan*. [http://hts.army.mil/Documents/Measuring%20Progress%20Afghanistan%20\(2\).pdf](http://hts.army.mil/Documents/Measuring%20Progress%20Afghanistan%20(2).pdf)
- [10] USAID. (2006). *Provincial Reconstruction Teams in Afghanistan: An Interagency Assessment*.
- [11] McChrystal, S. A. (2009). *Commander's Initial Assessment*. Cmdr NATO ISAF, Afghanistan.
- [12] Trochim, W. and Donnelly, J.P. (2007). *The Research Methods Knowledge Base*. 3rd edition. Thomson Publishing, Mason, OH.
- [13] Baker, T. (2010). *Final Interim Progress Report with MG Flynn: A Strategic Multilayer Assessment-Rich Contextual Understanding Project*. NSI Contract report.
- [14] Eles, P.T., Vincent E., Vasiliev B., & Banko K.M. (2012). *Opinion Polling in Support of the Canadian Mission in Kandahar – A Final Report for the Kandahar Province Opinion Polling Program, Including Program Overview, Lessons and Recommendations, and Results from the 2011 Polls*. (S/REL NATO/ISAF) DRDC CORA TR 2012 (under review).

REFERENCES

- [15] NATO Allied Command Operations Comprehensive Operations Planning Directive COPD Interim v1.0 (2010).
- [16] North Atlantic Treaty Organisation. NATO Operations Assessment Handbook, Interim version 1.0. (2011).
- [17] Department of Defense. (2010). Report on Progress Toward Security and Stability in Afghanistan. Report to Congress with section 1230 of the National Defense Authorization Act for Fiscal Year 2008.
- [18] Department of Defense. (2010). Assessment and Measures of Effectiveness in Stability Ops: Tactics, Techniques, and Procedures. Centre for Army Lessons Learned.
- [19] NATO SAS 091. NATO Operations Assessment: A Case Study Based on Planning for Transition in Afghanistan. September 2011.
- [20] Downes-Martin, S. (2010). Assessments Process for RC(SW). Centre for Irregular Warfare and Armed Groups. US Naval War College.
- [21] NATO RTO HFM 160. How to Improve your Aim: Measuring Effectiveness of Activities that Influence Attitudes and Behaviors. August 2011.
- [22] Doran, G. T. (1981). There's a S.M.A.R.T. Way to Write Management's Goals and Objectives. *Management Review*, 70, 35-36.
- [23] Marklund, J., Lindoff, J., Lackenbauer, H., & Harriman, D. (2011). Challenges in Assessing Progress in Multifunctional Operations: Experiences from a Provincial Reconstruction Team in Afghanistan. 16th International Command and Control Research and Technology Symposia (ICCRTS) (s. 22). Quebec City: The Command and Control Research Program (CCRP) within the Office of the Assistant Secretary of Defense (NII).
- [24] Hopkin, T., & Ramjeet, T. (2008). UK operational Analysis evaluation of current operational campaign effects assessment. DSTL/WP32694 v1.2, January 2009.
- [25] TNO. (2011). Lessons Identified and Best Practices NLD Participation in ISAF (URUZGAN) from 2006-2010. Workshop Proceedings.

Annex A – US INTERVIEW PROTOCOL

(Copied directly from the interview protocol administered by the US research team)

Overall instructions to the interviewer:

- *The interview is divided into a number of sections. Each section begins with a short explanation of the aim of the section (in italics) and when necessary some of the specific questions also have explanatory text.*
- *This information is directed to you, as an interviewer, to give you a better understanding of what type of information you should try to get from the interviewee.*
- *The questions are organised in order of priority and depending on how much time you have, you may skip some of the questions.*
- *The interviewer should note that not all questions will be applicable to every interviewee, as it depends on what position they held and what experiences they had.*
- *Furthermore, the general background information may not need to be asked in all cases, as this information can be obtained prior to the interview through either another interviewee or through research.*

Interviewer Name:

NOTE-TAKER:

Date:

Time: START: FINISH:

Interviewee Name:

Duration of MOST RECENT Posting TO JOINT Operation:

Location of OPERATION and Primary Posting:

Role/position/rank during posting:

Type of interview: COMMANDER or STAFF

Anonymity: Anonymous OR Attributable

ADDITIONAL ASSESSMENTS EXPERIENCE:

IRAQ

HAITI

BOSNIA

OTHER: _____

BEGIN COMMANDER & STAFF INTERVIEW

Green=Commander & Staff

Blue = Commander

Red = Staff

Hello and thank you for agreeing to participate in this study intended to improve and align assessments of progress in NATO joint operations.

[Interviewer assures that background items (above) are complete and asks for missing information. If the type of interview has not been determined ask the interviewee:

In your most recent NATO joint operation assignment did your primary duties include designing or producing assessments of mission progress?]

If “yes” use staff interview questions.]

Do you have any questions before we begin?

There are different kinds of assessments that can be conducted, on different levels and for different purposes. We would like to focus this interview on *comprehensive assessments of progress*, i.e., how you assess the progress of your comprehensive operations.

US Field Manual (FM) 3-07 – Stability Operations identifies a comprehensive approach to operations as one that “integrates the cooperative efforts of the departments and agencies of the United States Government, intergovernmental and nongovernmental organisations, multi-national partners, and private sector entities to achieve unity of effort toward a shared goal. “ – this reflects an operational environment with many blue side actors not pulled together to a single team, but acting cooperatively out of shared interest.

It is within the context of comprehensive and joint (NATO) operations that we would like to discuss your experience – broadly -- with assessments.

To make sure we are talking about the same thing, I want to read to how NATO defines an assessment:

“The purpose of Assessment is to support the NATO political domain and Commander and staff decision making in three areas:

- ... the progress of plan execution (actions / tasks).*
- ... the effectiveness of those executed actions by measuring the achievement of results (effects, objectives, and the end-state).*

- ... drawing conclusions about past situations, and in some cases mak[ing] forward-looking estimates about future trends, and recommend[ing] adjustments to the plan based on these conclusions.”]

1. [STAFF/COMMANDER] Would you consider that definition to be a generally accurate one for the types of comprehensive assessments you received and/or produced during your joint (NATO) operation experience ? Why or why not?

2. [COMMANDER] We are interested in the variety of (comprehensive) assessments you received or had access to during your most recent joint operations experience.

- 2a. Please identify the top 5 assessments rated in terms of your estimation of the **quality** of the information or analysis. Note that these need not be official or necessarily final documents or reports.

Intent of Question (2)

The aim of these questions is to provide an open ended opportunity to assure that the interview subject understands the intent of the questions. This includes determining what is meant by “comprehensive assessment” and to provide interviewers with an initial characterisation of the information they receive.

- 2b. For each can you tell me:
 - in what **format** it was it transmitted to you (e.g., message traffic, forwarded email, oral briefing, PP briefing, intel report, etc.)?
 - Whether the information/analysis timely? If applicable, can you give me a sense of the frequency with which you received the information or that the information was refreshed (e.g., monthly, quarterly, biannually)?
 - Which organisation produced the information or analysis (e.g. NATO, ISAF HQ, IJC/IDC, US DoD, other USG Agency, IC, other nation, etc.)?

3. [STAFF/SOME COMMANDERS] Think of the assessment product or type that you worked on the most. What was the process and who was involved in producing the assessments?

- 3a. Military and/or civilian?
- 3b. How many people or organisations were involved?
- 3c. If applicable, to which other staffs producing assessments was your team linked either formally or informally? Did your team have authority to task other assessment cells?
- 3d. What kind of resources did the assessment staff have, i.e. could they hire contractors?

- 3e. Who was the primary customer of the assessment?
- 3f. Did the customer generally dictate the contents of the assessment based on existing conditions or were there more static requirements?
- What was the frequency with which you produced or updated?
- How were assessments you produced typically presented to customers?
 - Visualisation methods (e.g., traffic lights)?
 - Text- and graphics-based reports?
 - Other?
 - Did you receive feedback on, or do you have an opinion about the ease of use for the customer of the method(s) you used?

Intent of Question (3)

The aim of this question is to get a better understanding of how assessment was actually conducted. Sub-questions in the 5th to 9th bullets are fairly technical and lower priority than the first four bullets.

4. [STAFF/COMMANDER] Is/was the assessment process formalised? (i.e. did you have a set format for assessments that you conducted or received? Did these templates come from higher commands or reachback cells, etc?)

- 4a. Did you have access to an assessment doctrine or guidelines?
- 4b. Are there dominant frameworks that are used to guide these assessments? If so, what are they?
- 4c. Was there any training for staff on how to produce the assessment and/or what data to use?
- 4d. Was the dissemination of this information strongly limited and if so, how/by whom??
- 4f. Did a specific entity maintain control over how the assessment information is interpreted and used?

Intent of Question (5):

To elicit more detail on the type of information/data provided to address each of six domains. Some current theories of comprehensive approach state that these five domains inter-relate and should all be addressed to reach a successful end state. Therefore, assessment of each of these domains should be conducted. The degree to which commanders express knowledge of all/some of these domains may indicate their appreciation of and interest in this concept and whether the comprehensive assessments received during their assignment were conducted with the required breadth. The expectation is that the subject will volunteer information on one or two domains and will likely need to be prompted for the others. There could be a large range of responses here in terms of granularity of information given, level of analysis, etc. Be prepared to probe in order to obtain the most usable information. Also, is the information provided of sufficient specificity (not too little or too much, not too simple or too complex) to make the effort of the comprehensive approach assessment worthwhile? Further, what was the impact of the assessment on the commander's view of the operational environment?

5. [STAFF/COMMANDER]

[For commander interviews the interviewer should ask the following. For staff interviews the interviewer should substitute the word "data" for "information" in the original question.]

We are interested in the types of assessment information you generally received, used or had access to and have identified six categories:

- (i) economic information about the host population;
- (ii) information about governance in the host country;
- (iii) rule of law among the citizens of the host country;

- (iv) social well being of the host population; or
- (v) the security conditions of the population
- (vi) the security conditions/situation of NATO or allied forces.

6a** Very briefly, what types of information (subjects or issues) do you associate with each of these domains?

[Prompt might be: For example what do you think of when you think of “social well being”?]

[For staff interviews substitute question below with:

“For each of the comprehensive assessments you produced please indicate which information domain(s) it primarily reported? Then skip to question 3c]

6b. For each of the top assessments you just listed (2a), please indicate whether it primarily reported:

- (i) economic information about the host population;
- (ii) information about governance in the host country;
- (iii) rule of law among the citizens of the host country;
- (iv) social well being of the host population; or
- (v) the security conditions of the population
- (vi) the security conditions/situation of NATO or allied forces.

[Interviewer identifies each assessment cited to prompt interviewee – if more than 5 cited, ask interviewee for the top 5 only, or to indicate a natural cut-off point in those mentioned on quality of information. Following questions should be asked for each assessment cited.]

6b1. Which information domains did it primarily report?

6b2. Did you find it specific to your needs?

6b3 Was the amount and kind of information useful overall (e.g., was it too technical, dumbed-down)?

6b4. Did the assessment contribute to your understanding of the operational environment? If so, how?

6b5.** Was this kind of information used as a supplement to reports of kinetic actions, or did it stand alone?

6b6. How well did the comprehensive assessments characterise progress towards strategic goals? ...operational goals? ...tactical goals?

Intent of Question (6b3)

Take other approaches to determining the usefulness of the comprehensive assessment at different levels.

- 6c.** Is there a category of information perhaps particular to your AOR or operational mission that is not covered by the six listed above? If yes, please indicate and explain a bit about it.

7. [STAFF] How were the data/information used in the assessment acquired and analysed?

Did you collect it in the field yourselves, use, polls, surveys, expert opinions, focus groups, media analysis?

- What were the main challenges in obtaining data?
- How were data aggregated and analysed (e.g., statistical analysis, trend analysis) were used?
- What timeframes of the plan (what time-scales are the results related to and how often is this timeframe revised)?
- Did you use any automatic aggregation, like traffic light systems?
- Were the data or the effects weighted, and if so, how?
- Were timeframes for the various expected effects of operations identified? By whom? Were these revised?
- What indicated “progress” (e.g., upward or downward statistically-significant trends; fewer bad things, more good things, etc.)? Who determined these and how?

8. We are interested as well in which sources and types of data you tended to find most credible/trustworthy and useful in assessing mission progress.

Thinking not just of assessments, but all sources of data you used (e.g., key leader engagements, ISAF, intel reports other civilian or military sources, SMEs) what were your most trusted and useful sources in each of the following domains?

[This is a double-barrelled question and interviewers should be sure to capture both trustworthiness and usefulness for each domain.]

- Economy
- Governance
- Rule of Law
- Social Well Being
- Security

Intent of Question (8)

This question should further clarify what information was influential in shaping the commander’s perspective and decision making. We are looking for an understanding of where the comprehensive approach assessments rank compared to other sources. This question could also provide insight into whether comprehensive assessments were ever used to resolve disputes between other sources of information (e.g., an intelligence report and a key leader engagement).

8a.** What kind of data, for example, qualitative data like interviews, anecdotes or quantitative data like statistics, did you tend to find most trustworthy and useful? Why?

9. Can you discuss any instances in which an assessment and a different source(s) gave contradictory information?

- 9a. If so, how often did this happen?
- 9b. Under what circumstances did this happen?
- 9c. Did this tend to occur in one information domain more frequently than in another?

- 9d. Did it tend to consistently involve one type of report or reporting source?
- 9e. What additional information did you request to help resolve the contradictions?
- 9f. How did you finally resolve these contradictions?

10. **[STAFF/ SOME COMMANDERS]** When you conducted your assessment of a mission, what plan did you assess against (e.g., an operational plan, a task-force plan, a strategic plan, or a diplomatic plan)?

- Did you assess against goals/objectives/effects articulated in that plan?
- Who owned this plan? (Is it a military, comprehensive or whole of government plan?)
- [STAFF] How did you identify what specifically to assess in terms of the goals/objectives in the plan?
- [STAFF] Were specific metrics/indicators/measures derived from the desired effects in the plan?
- [STAFF] If these metrics/indicators/measures needed to be modified for some reason (pragmatic to data collection, tailoring for environment, or otherwise), who had authority to do so and how did they go about making these decisions?
- [STAFF] What were the main challenges in formulating the assessment approach?
- Is the plan derived from or related to any other plans (i.e., NATO, ISAF, national, international)?
- Was the assessment staff part of the planning process? If so, how were they involved?

Intent of Question (10)

The intent of the question is to investigate whether the assessment cell has influence in the planning process and creation of effects.

11. Please explain whether and how comprehensive assessments ever influenced the plan and/or your activities?

- 11a. Were there changes of direction based on assessments that resulted in formal modifications to the plan?
- 11b. Were there changes in operational emphasis or activities as the result of assessments received?
- 11c. Did you reassign resources across activity domains as the result of assessments received?
- 11d. Can you give examples of each of these?

Intent of Question (11)

The intent of the question is to know if previous plans were continued throughout command rotations because ideally mission progress should have long term goals/objectives. We want to know to what extent commanders' intent changes over rotations.

12. **[STAFF]** How did assessment affect the plan?

- Did the assessments have any effect on the plan(s) (if so, what effect? If not, why not)? Can you give an example?

Intent of Question 12:

Since the purpose of assessment is to investigate if the plan works, it is important to investigate if the assessment actually had an impact on the current plan.

- What type of information from the assessment was typically fed back to the plan?
- At what time intervals did the feedback typically take place?

13. **[STAFF/SOME COMMANDERS]** Is the focus of the assessments affected by rotations in command?

- If so, how?
- Were there efforts to maintain continuity across command rotations? (e.g., were assessment data and/ or reports stored?)

Intent of Question 13:

We want to learn to what extent (and how) changes in rotation affect the quality of assessment, the day to day practice, and also if the same things are being assessed over time.

14.** Were you involved in hand-over of an AOR to a new commander (including yourself) during your deployment?

- **If no**, which assessments do you think would have benefitted new AOR commanders?
- **If yes** do you know which comprehensive assessments were made readily available and whether they were considered to be helpful and, how?
- Do you know whether any assessments were specifically requested and whether they were considered useful and how?

Intent of Question (14)

Identify what information was useful, highlighting the hand-over of an AOR to a new commander.

In hindsight, which assessments do you think would have benefitted new AOR commanders?

15. **[COMMANDER/STAFF]** A lot of issues have been discussed, and I want to make sure that we've captured the key points:

Based on your years of experience, in general what would you consider to be a useful 'assessment' of progress? What kind of content would such an 'assessment' cover? Which categories of information would it entail?

Intent of Question (15)

The interviewer should sum up the interview, make sure that all key points were captured, thank the participant for their time and tell them that they will be sent the notes.

15a. Can you summarise the three main challenges to conducting comprehensive assessments and, if possible, how they could be addressed?

15b. Can you summarise the top three current best practices or good ideas that you think should be incorporated into the practice of assessment (what are the things that really work or really could work well?)

15c. Can you summarise the top three types of additional assessments or data sources you would have liked to have during your assignment? Why were these data sources unavailable to you?

15d Finally, are there any concluding remarks or anything in general that you would like to add?

Thanks very much for your time! We will ensure that you receive a copy of the notes summarising our discussion today. You will then be able to comment on these notes to ensure that we captured your points accurately.



REPORT DOCUMENTATION PAGE			
1. Recipient's Reference	2. Originator's References STO-TR-HFM-185 AC/323(HFM-185)TP/489	3. Further Reference ISBN 978-92-837-0180-4	4. Security Classification of Document UNCLASSIFIED/ UNLIMITED
5. Originator Science and Technology Organization North Atlantic Treaty Organization BP 25, F-92201 Neuilly-sur-Seine Cedex, France			
6. Title Processes for Assessing Outcomes of Multi-national Missions			
7. Presented at/Sponsored by Findings of Task Group HFM-185.			
8. Author(s)/Editor(s) Multiple			9. Date November 2013
10. Author's/Editor's Address Multiple			11. Pages 78
12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other STO unclassified publications is given on the back cover.			
13. Keywords/Descriptors <div style="display: flex; justify-content: space-between;"> <div> Mission Assessment Campaign Assessment Measures of Effect </div> <div> Measures of Performance Metrics Cross-mission Comparison </div> </div>			
14. Abstract Multi-national missions require comprehensive assessment that measures the impact of military and civilian actions on opposing forces and the local population. In order to identify and characterize limitations of current assessments and identify best practices, this NATO RTG task group examined current measurement and assessment practices by NATO members in current and recent theatres of operations. The task group also conducted a preliminary study to identify a potential set of core indicators that could provide a high-level assessment of macro change over time to compliment campaign assessments aimed at specific interventions. These core indicators, if used to assess all missions, would allow for cross-mission comparison, a capability that currently does not exist. Based on the findings, the task group recommends that NATO develop the policies, plans, infrastructure and procedures to successfully conduct comprehensive assessments. In order to do this NATO must develop standard assessment procedures and frameworks for use throughout NATO/ISAF and other multi-national operations. This includes common lexicons and definitions used by member nations and a core set of metrics to supplement campaign measures of performance (MOPs) and measures of effectiveness (MOEs) and allow for comparisons between NATO missions.			





BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cs0.nato.int**DIFFUSION DES PUBLICATIONS****STO NON CLASSIFIEES**

Les publications de l'AGARD, de la RTO et de la STO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la STO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la STO, de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre et la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la STO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (<http://www.sto.nato.int/>) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX**ALLEMAGNE**

Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7, D-53229 Bonn

BELGIQUE

Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research
for Defence, National STO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Bruxelles

CANADA

DSIGRD2 – Bibliothécaire des ressources du savoir
R et D pour la défense Canada
Ministère de la Défense nationale
305, rue Rideau, 9e étage
Ottawa, Ontario K1A 0K2

DANEMARK

Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESPAGNE

SDG TECIN / DGAM
C/ Arturo Soria 289
Madrid 28033

ESTONIE

Estonian Ministry of Defence
Estonian National Coordinator for NATO STO
Sakala 1
Tallinn 15094

ETATS-UNIS

Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc - BP 72
92322 Châtillon Cedex

GRECE (Correspondant)

Defence Industry & Research General
Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE

Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25, H-1885 Budapest

ITALIE

Centro Gestione Conoscenza
Secretariat General of Defence
National Armaments Directorate
Via XX Settembre 123/A, 00187 Roma

LUXEMBOURG

Voir Belgique

NORVEGE

Norwegian Defence Research
Establishment, Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

PAYS-BAS

Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE

Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL

Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide, P-2720 Amadora

REPUBLIQUE TCHEQUE

Vojenský technický ústav s.p.
CZ Distribution Information
Mladoboleslavská 944
PO Box 18
197 06 Praha 9

ROUMANIE

Romanian National Distribution
Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353 Bucharest

ROYAUME-UNI

Dstl Knowledge and Information
Services
Building 247
Porton Down, Salisbury SP4 0JQ

SLOVAQUIE

Akadémia ozbrojených síl gen.
M.R. Štefánika, Distribučné a
informačné stredisko STO
Demánová 393
031 06 Liptovský Mikuláš 6

SLOVENIE

Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

TURQUIE

Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi
Başkanlığı
06650 Bakanlıklar – Ankara

AGENCES DE VENTE

**The British Library Document
Supply Centre**
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

**Canada Institute for Scientific and
Technical Information (CISTI)**
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2
CANADA

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications STO, RTO et AGARD figurent dans le « NTIS Publications Database » (<http://www.ntis.gov>).



BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cs.o.nato.int



DISTRIBUTION OF UNCLASSIFIED STO PUBLICATIONS

AGARD, RTO & STO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all STO reports, or just those relating to one or more specific STO Panels, they may be willing to include you (or your Organisation) in their distribution.

STO, RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for STO, RTO or AGARD documents should include the word 'STO', 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of STO reports as they are published, please visit our website (<http://www.sto.nato.int/>) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM

Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research
for Defence, National STO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30
1000 Brussels

CANADA

DRDKIM2 – Knowledge Resources Librarian
Defence R&D Canada
Department of National Defence
305 Rideau Street, 9th Floor
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC

Vojenský technický ústav s.p.
CZ Distribution Information Centre
Mladoboleslavská 944
PO Box 18
197 06 Praha 9

DENMARK

Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESTONIA

Estonian Ministry of Defence
Estonian National Coordinator for NATO STO
Sakala 1, Tallinn 15094

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc - BP 72
92322 Châtillon Cedex

GERMANY

Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7
D-53229 Bonn

GRECE (Point of Contact)

Defence Industry & Research General
Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HUNGARY

Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25, H-1885 Budapest

ITALY

Centro Gestione Conoscenza
Secretariat General of Defence
National Armaments Directorate
Via XX Settembre 123/A, 00187 Roma

LUXEMBOURG

See Belgium

NETHERLANDS

Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY

Norwegian Defence Research
Establishment, Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

POLAND

Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL

Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide, P-2720 Amadora

ROMANIA

Romanian National Distribution
Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6, 061353 Bucharest

SLOVAKIA

Akadémia ozbrojených síl gen
M.R. Štefánika, Distribučné a
informačné stredisko STO
Demänová 393
031 06 Liptovský Mikuláš 6

SLOVENIA

Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

SPAIN

SDG TECIN / DGAM
C/ Arturo Soria 289
Madrid 28033

TURKEY

Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi
Başkanlığı
06650 Bakanlıklar – Ankara

UNITED KINGDOM

Dstl Knowledge and Information
Services
Building 247
Porton Down, Salisbury SP4 0JQ

UNITED STATES

Defense Technical Information
Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

SALES AGENCIES

**The British Library Document
Supply Centre**
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

**Canada Institute for Scientific and
Technical Information (CISTI)**
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2
CANADA

Requests for STO, RTO or AGARD documents should include the word 'STO', 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of STO, RTO and AGARD publications are given in "NTIS Publications Database" (<http://www.ntis.gov>).